SCIENTIFIC REPORTS
nature research

OPEN

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

Joanne C. Van Slooten[1*], Sara Jahfari[2,3,4] & Jan Theeuwes[1,4]

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.

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[1]. The neuromodulator dopamine (DA) plays an important role in how individuals learn from their interactions with the environment[2,3] and has also been linked to individual variability in spontaneous eye blink rate (sEBR)[4–6]. While research suggest that sEBR reflects the extent to which individuals learn from negative outcomes of their actions[5], 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[7]). In general, pharmacological studies in animals and humans have shown that DA-enhancing drugs elevate sEBR, while DA-decreasing drugs suppress them[4,6,8–12]. Moreover, sEBR is altered in clinical conditions that are associated with dysfunctions of the DAergic system[13,14]. For example, sEBR is decreased in Parkinson's disease (PD)[15,16], 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[17] in proportion to the post-mortem measured DA concentrations in the caudate nucleus[18]. 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[19,20].

[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. [4]These authors jointly supervised this work: Sara Jahfari and Jan Theeuwes. *email: joannevslooten@gmail.com

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[5,6] and reversal learning[21]. In particular, two studies found that sEBR predicted RL effects on future value-based choices[5,6]. 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[1]: 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. On the one hand, avoiding the least-rewarded option in the test phase could be caused by enhanced learning from negative outcomes (negative learning rate). On the other hand, by an exploitative choice strategy (explore-exploit tendency) in which the most-rewarded option is consistently chosen, hence, the least-rewarded choice option is learned to be avoided[22]. This makes previous findings[5] ambiguous regarding which specific cognitive processes sEBR reflects. Even more so as recent literature suggests very different dopaminergic mechanisms in using value to make decisions (explore-exploit) and updating values (learning)[23–25].

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 (Fig. 1a). To specify these underlying processes, we used a hierarchical Bayesian version of the Q-learning RL model[22,26,27] (Supplementary Fig. 1). 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 (β; Fig. 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[28] (α; Fig. 1b). As value beliefs are differently updated after positive and negative reward prediction errors via striatal D1 and D2 receptors[29], we defined separate learning rate parameters for positive (α_{Gain}) and negative (α_{Loss}) feedback[22,27,30–33].

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[34] (Fig. 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 (β), but not the tendency to learn from negative feedback (α_{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.

Results

Blinking. On average, participants blinked 12 times per minute (median = 10.6; SD = 8.3, range = 1.3–34.9; Fig. 2a), a rate that is comparable to earlier reports[5,35,36]. 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 ($t(19.8) = 1.26$, $P = 0.22$, Welch's t-test; $BF_{10} = 0.61$).

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 ($r = -0.46$, $P = 0.005$; $BF_{10} = 12.52$, Fig. 2b). As shown in Fig. 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 ($F(1,34) = 7.23$, $P = 0.01$; $BF_{10} = 17.6$), and a trend towards an interaction effect ($F(2,68) = 2.66$, $P = 0.08$). This was consistent with a Bayesian Mixed ANOVA revealing that the interaction effect model was only slightly preferred over the main effect model by a BF of 1.04. Exploratory post-hoc tests suggested that lower sEBR related to better learning performance in the more certain AB ($t(34) = -2.5$, $P = 0.02$; $BF_{10} = 3.18$) and CD pairs ($t(34) = -3.7$, $P < 0.001$; $BF_{10} = 39.4$), but not in the uncertain EF pair ($t(34) = -0.5$, $P = 0.59$, $BF_{10} = 0.36$).

In the transfer phase, all participants were able to approach the most rewarded option (approach-A: mean accuracy = 80%, SD = 24%) and to avoid the least rewarded option (avoid-B: mean accuracy = 79%, SD = 21%) well above chance (one-sample t-test; both P-values < 0.001), 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 neither the observed interaction ($F(1,34) = 1.79$, $P = 0.2$; BF = 5.5 in favour

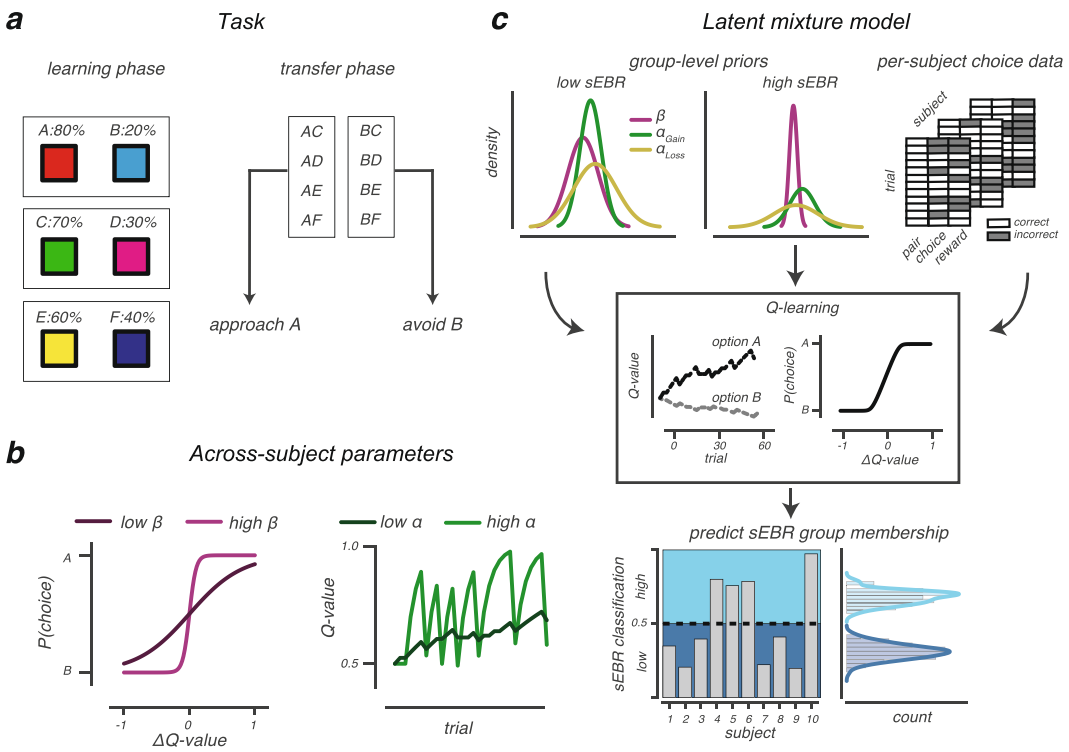

Figure 1. Task and model. (**a**) In the learning phase (left), three different option pairs (AB, CD and EF) were presented in random order and participants had to learn to select the most rewarding option of each pair (A, C and E). Each choice was followed by probabilistic auditory feedback indicating 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 six options were paired with one another to create 12 novel options, 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[5]. (**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 exploitatory 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 the low or high sEBR group, given observed learning.

of the null-model), nor the correlation between sEBR and avoid-B accuracy ($r = -0.29$, $P = 0.08$, $BF_{10} = 0.88$) reached significance (Fig. 2d).

As fatigue is tied to poorer task performance and increased blink rates and blink durations[37–41], 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 ($r = 0.18$, $P = 0.28$, $BF_{10} = 0.36$), nor with sEBR ($r = 0.04$, $P = .8$, $BF_{10} = 0.21$). Additional analyses of learning phase choice reaction times and sEBR showed no relation ($r = 0.22$, $P = 0.2$, $BF_{10} = 0.46$), indicating sEBR did not predict differences in selection speed. Based on these results, we did not find evidence that performance differences during learning between sEBR groups were explained by differences in fatigue.

To summarise, our behavioural results suggest that individual variability in sEBR relates to how participants learn from probabilistic feedback, with lower sEBR predicting better learning, especially from more reliable feedback.

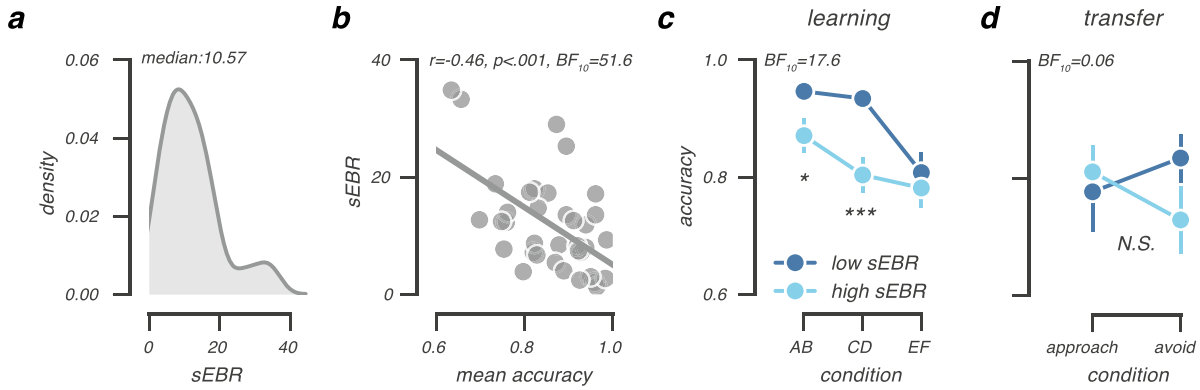

Figure 2. sEBR data and choice performance in the learning and transfer phase. (**a**) sEBR distribution across participants (N = 36), recorded prior to the probabilistic RL task. (**b**) Lower sEBR predicts better overall choice performance in the learning phase. This correlation was explained by higher choice accuracy in the AB and CD pairs, but not in the EF pair (**c**). (**d**) In the transfer phase, choice performance was nummerically comparable to previous research[5], 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 < 0.05; **P < 0.001; BF_{10} = evidence in favour of the alternative model.

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

We first assessed the predictive accuracy of our model by performing parameter recovery on the α_{Gain}, α_{Loss} and β-parameter (see also *Methods*). This procedure evaluates whether the fitted model produces meaningful parameter values in a scenario where data is generated (simulated) using the originally estimated parameter values[42]. As shown in Fig. 3a, true (estimated) and recovered (simulated) parameter estimates were tightly correlated across all three parameters ($r > 0.99$; $P < 0.001$; $BF_{10} = \infty$), indicating that the parameters were well recovered by our model. Second, we used posterior predictive checks (PPC) on learning curves of the AB, CD and EF pair to evaluate whether our model could reproduce participants' choice behaviour in the learning phase. As can be seen in Fig. 3b, our model correctly captured learning curves across all three learning pairs and separate sEBR groups. Finally, we evaluated a simpler Q-learning model with a single learning rate that was agnostic to the sign of the reward prediction error and used model comparisons to show that a model with two learning provided the best fit to the data (see *Methods*). Specifically, model comparison using Pareto smoothed importance-sampling leave-one-out cross-validation (PSIS-LOO) indicated the model with two learning rates best described choices in the learning phase (elpd difference = 289.23; SD = 51.98). Together, these analyses suggest our model provides a good description of choice behaviour in the learning phase.

Next, we evaluated the relationship between sEBR and the estimated Q-learning model parameters to understand how sEBR related to learning. As shown in Fig. 4, we observed shifts between the high and low sEBR groups in the group-level posterior distributions of all parameters, but particularly for the β- and α_{Loss}-parameter. These observations suggested that the low sEBR group exploited high value options more often (higher β-parameter) and updated value beliefs stronger after negative feedback (higher α_{Loss}-parameter). Note, however, that these observations were based on visual inspections of the group-level posteriors. To formally test whether high and low sEBR groups can be distinguished on the basis of the observed differences in the estimated Q-learning parameters, we used a recently developed Bayesian latent mixture modelling approach[43] that we adapted for Q-learning (Fig. 1c).

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

As shown in Fig. 5a, our Bayesian latent mixture model correctly classified 72% of participants using the estimated Q-learning parameters, a percentage that was well above chance ($P = 0.011$, $BF_{10} = 14.5$; *one-sided binomial test*). Consistently, higher probabilities to be classified as a member of the high sEBR group by the latent mixture model predicted higher sEBR values ($r = 0.51$, $P < 0.001$, $BF_{10} = 24.9$, Fig. 5b), 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

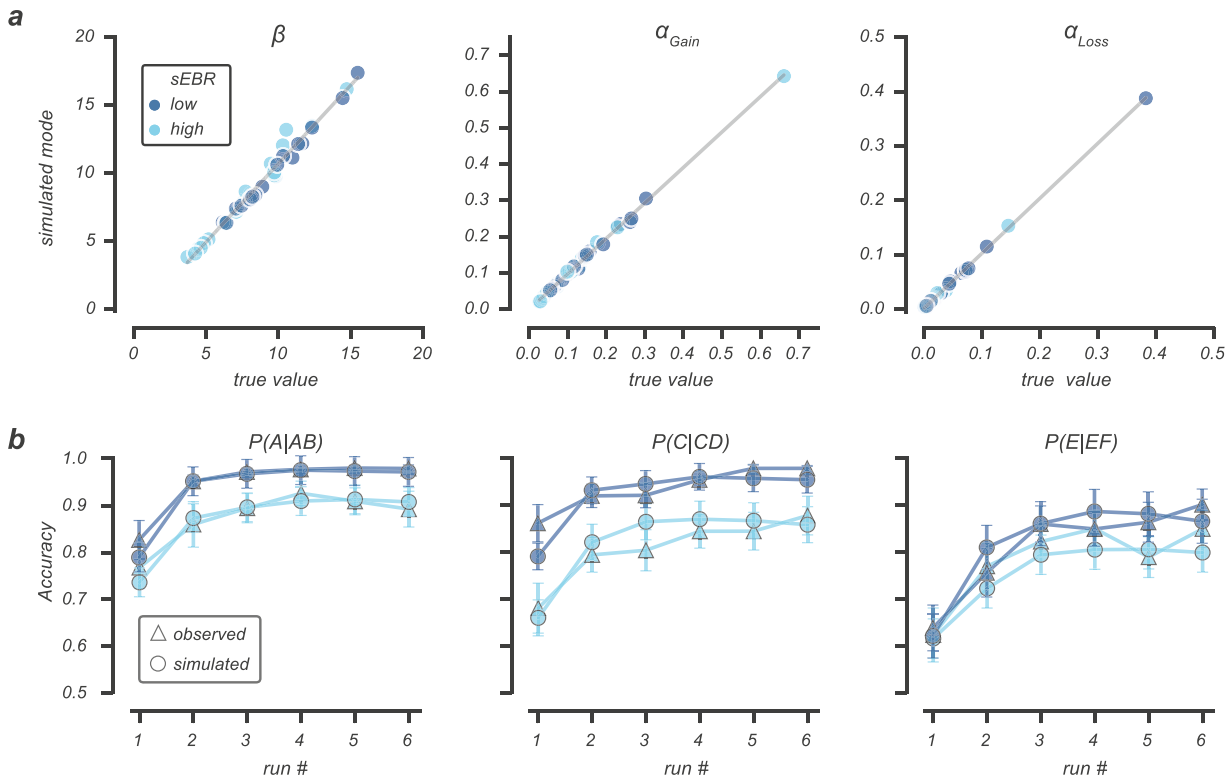

Figure 3. Q-learning parameter recovery and posterior predictive checks for high and low sEBR groups. (**a**) The close correspondence between each participant's true (observed) parameter modes and simulated modes indicate the Q-learning model is well able to recover the original parameters that were used for data simulations. (**b**) Participants' choice accuracy averaged across six bins of 60 trials (observed; triangle markers) was plotted against simulated data (simulated; circle markers) by using parameter draws from the posterior predictive distribution. Shown separately for the two sEBR groups (low, high) and different option pairs (AB, CD, EF), the Q-learning model correctly predicts participants' observed choice patterns. Error bars represent SEM.

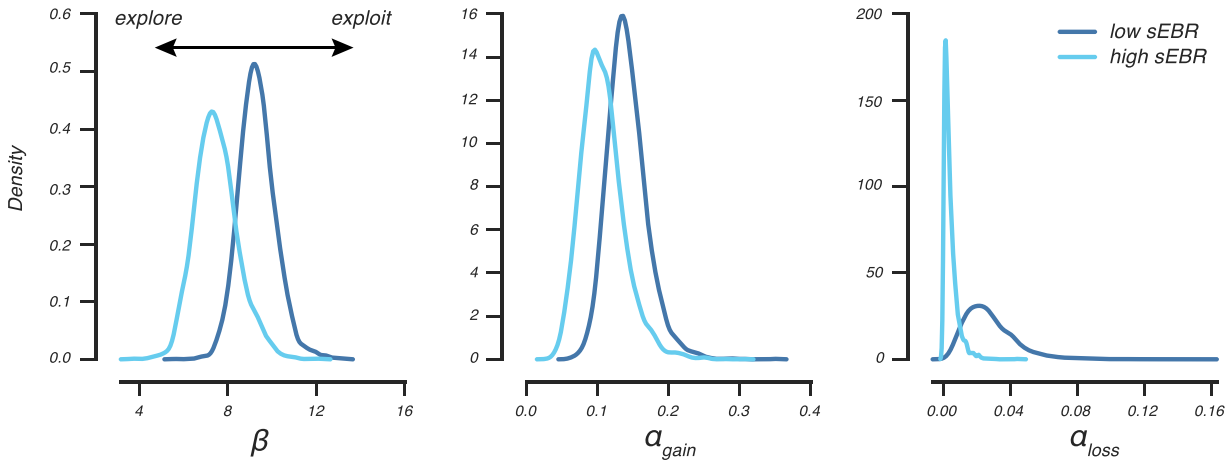

Figure 4. Q-learning parameter estimation for low and high sEBR groups. Posterior distributions of group-level parameters for high and low sEBR groups obtained by fitting the Bayesian hierarchical Q-learning model separately for both groups.

question how sEBR relates to individual variability in how we update our beliefs after desired (α_{Gain}) and undesired (α_{Loss}) outcomes, or the variability by which we exploit actions that will likely result in reward (β).

We used a multiple regression model that incorporated all three cognitive processes (α_{Gain}, α_{Loss} and β) to explain individual variability in sEBR. The model well accounted for the variability in sEBR ($F_{(3,32)} = 5.8$, $P = 0.003$, $R^2 = 0.35$), which was driven by a significant contribution of the β-parameter (b_β (SE) $= -4.5$ (1.2), $z = -3.7$, $P < 0.001$, $BF_{10} = 33.8$), but not the α_{Gain}- ($b_{\alpha_{Gain}}$ (SE) $= -1.5$ (1.4), $z = -1.1$, $P = 0.28$, $BF_{10} = 1.2$) or the

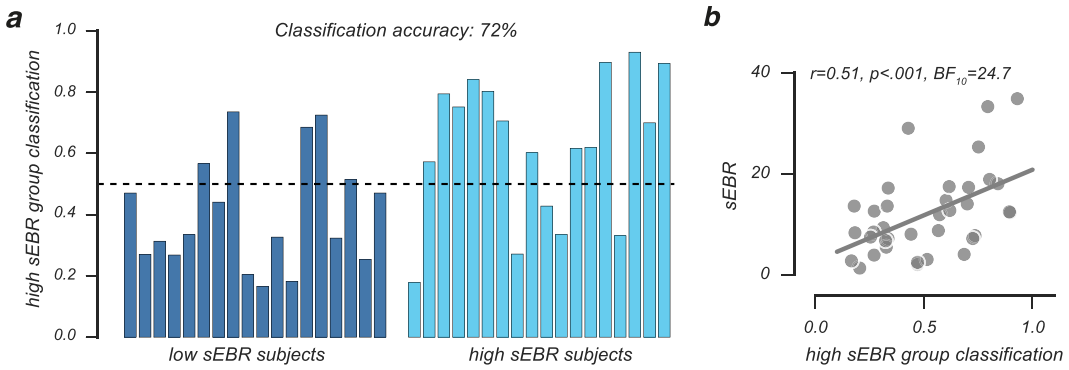

Figure 5. Bayesian latent mixture model classification of sEBR group membership. (**a**) Per-participant posterior classification probability to belong to the high sEBR group. A low posterior classification probability suggest that a participant is very likely to fall into the low sEBR group, whereas 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.

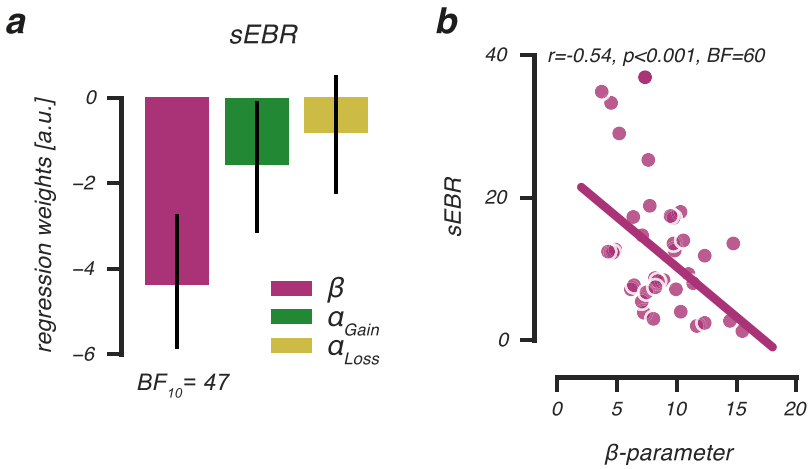

Figure 6. sEBR predicts individual differences in exploration and exploitation. (**a**) Beta coefficients of a multiple regression analysis indicating that β-parameter estimates uniquely and negatively relate to sEBR. This was further illustrated by a negative correlation between individual β-parameter estimates and sEBR (**b**), showing that low sEBR individuals exploited highly values options more often compared to high sEBR individuals. Error bars represent SEM.

α_{Loss}-parameter ($b_{\alpha_{Loss}}$(SE) = −0.5 (1.5), $z = -0.4$, $P = 0.71$, $BF_{10} = 0.8$). As shown in Fig. 6a, the Bayesian linear regression analysis further indicated that the model that only incorporated the β-parameter to explain individual variability in the sEBR data was 47 times more likely to explain the data compared to the null-model, which is regarded very strong evidence in favour of this model[44] (Supplementary Table 1). Figure 6b illustrates the negative relationship between the β-parameter and sEBR, indicating that exploitative decision makers had a lower sEBR. Together, these results link sEBR to individual variability in exploiting actions that lead to rewarding outcomes, but not to the magnitude by which individuals update their value beliefs after positive or negative outcomes.

Learning effects on choices in the transfer phase. Our results thus far relate sEBR to how participants make value-based choices during learning, but show no reliable effect of sEBR on avoid-B or approach-A choices in the transfer phase. Because this relationship has been reported in the past[5], this section additionally examined how the Q-learning parameters (α_{Gain}, α_{Loss}, β), 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 $F_{(3,32)} = 3.7$, $P = 0.02$, $R^2 = 0.26$; Fig. 7). This was driven by a significant contribution of the β-parameter (b_β(SE) = 0.069 (0.03), $z = 2.066$, $P = 0.047$, $BF_{10} = 2.4$) and a smaller, albeit non-significant, contribution of the α_{Loss}-parameter ($b_{\alpha_{Loss}}$(SE) = 0.072 (0.04), $z = 1.8$, $P = 0.08$, $BF_{10} = 2.4$). The Bayesian linear regression analysis further indicated that a model that incorporated both the β- and α_{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

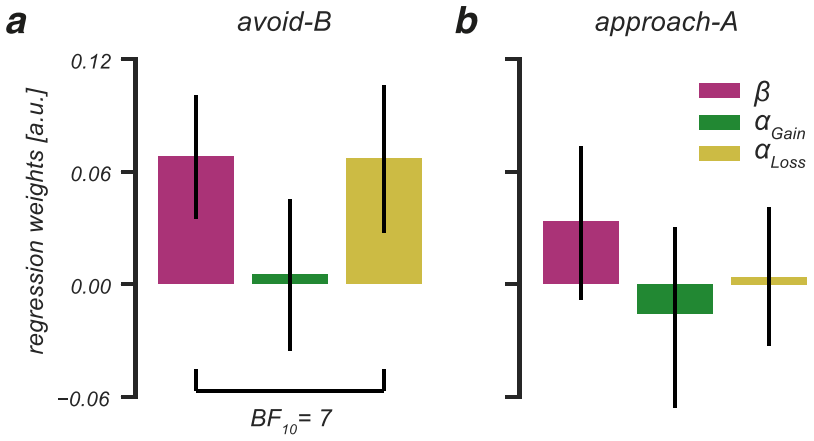

Figure 7. Avoid-B, but not approach-A, choices in the transfer phase are related to individual variability in negative learning rates and explore-exploit tendencies during learning. (**a**) Exploitation of high valued options (high β) and enhanced learning from negative feedback (high α_{Loss}) during 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. Error bars represent SEM.

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.

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 β (which also predicted sEBR during learning) and α_{Loss} (which is hypothesized to be associated with variability in sEBR[5]). 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 variables after controlling for all other information included[45]. Supplementary Table 3 shows all partial correlations between the variables, which are graphically depicted in Fig. 8. In this graph, three important between-node relationships were observed. First, individual differences in sEBR were significantly and negatively related to the β-parameter (*partial* $r = -0.515$, $P < 0.001$), consistent with our previous finding that exploitative decision makers had a lower sEBR. Second, the α_{Gain}- and α_{Loss}-parameter were significantly and positively related to each other (*partial* $r = 0.522$, $P < 0.001$), but not to sEBR, which is inconsistent with earlier work that hypothesized sEBR indexes how much individuals learned from the negative outcomes of their choices[5]. Lastly, the ability to avoid the least rewarded option in the transfer phase related to the β- and α_{Loss}-parameter, consistent with our previous results. However, the network analysis indicated these relationships were not robust. More importantly, the ability to avoid the least rewarded option was unrelated to sEBR, an observation that is not in line with earlier work[5]. Overall, this analysis paints a clear picture of how sEBR relates to learning and subsequent value-based choices, namely that it uniquely reflects a decision maker's explore-exploit tendency during learning.

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 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 studies investigating cognitive flexibility, which is considered a behavioural component of explore-exploit decision-making[24]. 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[21], working memory[46] and attentional set-shifting tasks[36,47–49]. As exploration or enhanced cognitive flexibility supports behaviour aimed at detecting novel information, this either improves or deteriorates

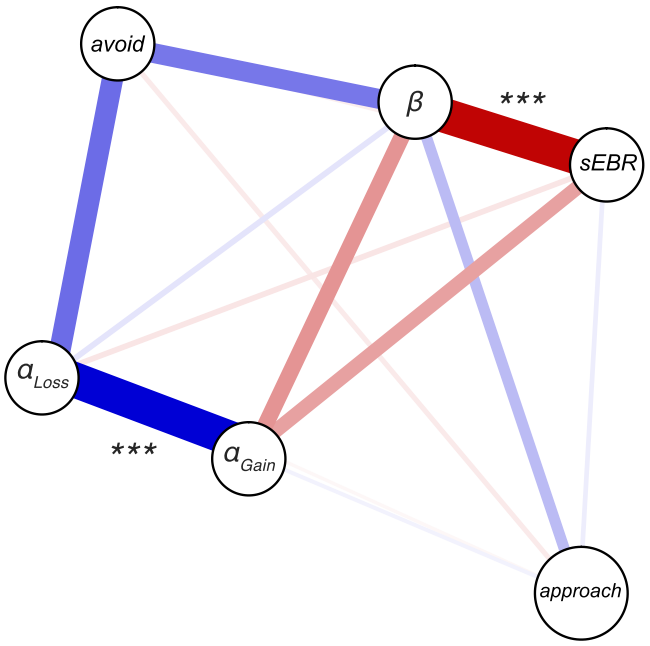

Figure 8. Network analysis. Graphical depiction of the partial correlation network of sEBR, approach-A/avoid-B performance and the cognitive processes underlying learning. Variables of interest are represented as nodes. The estimated relations between variables are represented as edges, where the colour 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.

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[22]. This explains why exploitative individuals with a relatively low 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 most rewarded option.

Previous work that investigated sEBR in the context of probabilistic RL hypothesised that sEBR predicted how much individuals learned from the negative outcomes of their choices during prior learning[5]. 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 (α_{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 individuals learn from feedback, we find that it can be used to index how individuals use value to make a value-based choice.

While we observed strong effects of sEBR during learning, effects on later value-based choices were rather weak or unreliable. Our network analysis - in which the unique relationship between two variables was estimated after controlling for the influence of all other variables - indicated that sEBR was unrelated to participants' ability to avoid the least rewarded option. This finding is inconsistent with earlier work observing that sEBR did relate to the ability to avoid the least rewarded option[5], or that it predicted the modulatory effect of dopaminergic drugs on approach and avoidance behaviours[6]. An important difference between these and our study, is that we evaluated both the effects of sEBR on learning as well as on later value-based choices, in one model. This analysis indicated that sEBR primarily related to an individual's explore-exploit tendency during learning, that *in turn* related to the ability to avoid the least rewarded option in the transfer phase. Thus, individuals with a lower sEBR tended to exploit high valued outcomes, which especially improved learning in the AB pair and might be the reason that they avoided the least rewarded option B in the transfer phase. This could suggest that earlier observed effects of sEBR on approach and avoidance behaviours may be driven by underlying variability in explore-exploit tendencies. Nevertheless, various studies have shown that separate "Go" (approach-A) and "NoGo" neuronal populations (e.g. avoid-B) represent positive and negative action values that determine action selection[50–52]. Future

studies that include dopaminergic manipulations combined with computational modelling to evaluate how sEBR relates to learning and later value-based decision biases might provide fruitful to answer this question.

Our observation that sEBR primarily reflects individual explore-exploit tendencies during learning could reconcile our work with the aforementioned studies[5,6], as these and other studies[4,8,18] have suggested that sEBR may reflect tonic, or baseline, striatal dopamine levels. It has been proposed that fluctuations in tonic dopamine levels predominantly affect the expression, rather than learning, of motivated behaviour[53,54], which agrees with our finding that sEBR uniquely predicted how value was used to make decisions. For example, studies have shown that mice with chronically elevated tonic DA levels were highly motivated to work for food rewards, even when their increased efforts did not result in better outcomes[55–57]. Conversely, depleted tonic DA levels in nucleus accumbens lowered motivation to work for rewards[58]. These findings agree with computational modelling studies observing that genetic, simulated or pharmacological differences in tonic DA levels uniquely related to explore-exploit tendencies, but not to learning rates[57,59–62]. Also in Parkinson's patients, some effects of dopaminergic medication on reward and punishment learning can be explained by motivational differences at the time of choice, rather than by differences in feedback learning[63–65]. Together, these studies suggest that tonic DA levels impact the expression of motivated behaviour, or more specifically, explore-exploit tendencies. With respect to our findings, higher sEBR - potentially indexing higher tonic DA levels - may reflect increased motivation and energy expenditure to promote the exploration of novel options. Lower sEBR - potentially indexing lower tonic DA levels - may reflect decreased motivation and energy conservation to promote the exploitation of options with known reward. We note, however, that in a recent study where Parkinson patients were evaluated on and off medication with a similar task and model, we observed no reliable within-patient changes in the explore-exploit trade-off[66]. Thus, on the behavioural level our data agree with studies linking sEBR to tonic DA levels and variability in explore-exploit tendencies. However, our data preclude any strong conclusions about the biological mechanisms affecting sEBR without any direct manipulations of DA, which should be the focus of future studies.

To conclude, sEBR predicted an individual's explore-exploit tendency during learning, thereby reflecting the sensitivity to value differences during a value-based choice. To our knowledge, this study is the first to associate sEBR to the underlying cognitive processes of learning, thereby providing a mechanistic understanding of the relation between sEBR, learning and the effects of learning on future value-based choices. We believe that using these methods advances our understanding of how sEBR relates to DA-dependent cognitive performance which may unify the diverse behavioural effects linked to sEBR, such as punishment or avoidance learning[5,6], reversal learning[21], as well as cognitive flexibility[36,47–49]. Together, our results indicate that sEBR can be used as an easy to measure behavioural index of individual explore-exploit tendencies during learning. Whether this is driven by fluctuations in tonic DA levels should be validated by other studies that directly measure or manipulate DA in a reinforcement learning task design.

Methods

Participants. The pupillometry data of the current data set was previously published[27], 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. Note that the current dataset includes four more participants compared to the previously published one[27] where these participants were excluded due to inadequate fixation to the centre of the screen during reinforcement learning which rendered their pupil data unreliable.

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 1000 Hz seven minutes of spontaneous eye blinks from the continuously tracked eye data, which provides reliable sEBR estimates[67]. 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[68]. 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[69] that consisted of a learning and a transfer phase. For an extended description of the task, stimuli and trial structure, we refer to[27]. 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 group-level median, leaving 36 participants for subsequent analyses. All 36 participants reliably choose A over B in the test phase; a learning criterion that has previously been used in the context of this task[5,69]. 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[1,70] 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:

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

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 prediction errors on current value beliefs. A relatively high learning rate indicates more sensitivity to recent prediction errors, whereas a relatively low learning rate indicates a stronger focus on the integration of prediction errors over multiple trials[30]. Modeling two learning rates was validated by comparing this model to a hierarchical Q-learning model with a single learning rate that was agnostic to the sign of the reward prediction error. Model comparison was based on Pareto smoothed importance-sampling leave-one-out cross-validation (PSIS-LOO)[71] that uses the difference in the estimated log predictive density (elpd) between the two models to evaluate differences in model fit. This analysis showed a positive elpd difference (elpd diff = 289.23), that was larger than the estimated standard error (SD = 51.98), indicating the model with two learning rates had higher prediction accuracy compared to the one with a single learning rate. This finding was further highlighted by model performance evaluations using posterior predictive checks (Supplementary Fig. 2) and agrees with other studies showing superior performance of a Q-learning model with separate learning rates to explain choice behaviour in probabilistic selection tasks[26,27,30,31,33,66,72].

A choice between two presented stimuli on the next trial was described by a "softmax" choice rule:

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

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 β value indicate greater sensitivity to smaller value differences, hence, exploitative choices for high reward options (Fig. 1b).

Bayesian hierarchical implementation of the Q-learning model. We implemented the Q-learning model in a hierarchical Bayesian framework (Supplementary Fig. 1)[22,26,27,73], 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[74,75]. 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 (α_{Loss}, α_{Gain}) and explore-exploit tendencies (β). For an extended description of the applied Bayesian hierarchical model, we refer to[27].

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 (α_{Loss}, α_{Gain} and β) underlying learning (Fig. 1c)[34,76]. 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 or high sEBR group, respectively. For each participant, the posterior mean of the x_i variable 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 (α_{Gain}, α_{Loss} and β) for the low and high sEBR groups. Specifically, for each group probit transformed individual-level parameters were drawn from 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 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[34]. 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 and validation. Our model-based analyses were implemented in PyStan mc-stan.org and fit to all trials of the learning phase that fell within the correct response time window 150 ms $\leq$ RT $\leq$ 3500 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[74], a convergence diagnostic that compares between and within chain variability, as all Rhats were $<$1.05. We further assessed the predictive accuracy of our Bayesian hierarchical Q-learning model, by performing parameter recovery and posterior predictive checks (Fig. 3a). For parameter recovery, we selected the mode of the posterior parameter distributions of each participant to simulate 200 new learning phase datasets per participant. The originally observed parameter estimates (α_{Gain}, α_{Loss} and β) were correlated with the parameter modes of the 200 simulation fits to evaluate our model's ability to recover the originally observed parameter estimates used for the simulations. Posterior predictive checks were calculated for mean choice accuracy across the learning phase by sampling 500 parameter sets from the joint posterior distribution and generating 500 independent learning phase datasets using those parameters. From these datasets mean accuracy was calculated for each dataset separately for learning pairs and trial bins (Fig. 3b).

Multiple regression analyses. We performed frequentist and Bayesian multiple regression analyses in JASP jasp-stats.org to quantify the relative influence of each model parameter (α_{Gain}, α_{Loss} and β) 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 evaluated step by step. Resulting Bayes Factors (BF) are interpreted as the odds supporting one model over another. BF-values between 3–10 indicate substantial support for the alternative model over the null model that a regressor's true value is zero, whereas BF-values $>$ 10 indicate strong support that the alternative model is favoured over the null model[44]. For all analyses, we selected the modes of the individual posterior parameter distributions of all participants. These variables were log-transformed and normalised prior to analysis to account for parameter skewness and scaling effects.

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

Data availability

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

Received: 10 June 2019; Accepted: 31 October 2019;
Published online: 22 November 2019

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 function: A 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**, 503–507 (1982).
17. Lawrence, M. & Redmond, D. Jr. MPTP Lesions and Dopaminergic Drugs Alter Eye Blink Ratein African Green Monkeys. *Pharmacology Biochemistry & Behavior* **38**, 869–874 (1991).
18. Taylor, J. R. *et al*. Spontaneous Blink Rates Correlate with Dopamine Levels in the Caudate Nucleus of MPTP-Treated Monkeys. *Experimental Neurology* **158**, 214–220 (1999).
19. Sescousse, G. *et al*. Spontaneous eye blink rate and dopamine synthesis capacity: preliminary evidence for an absence of positive correlation. *European Journal of Neuroscience* **47**, 1081–1086 (2018).
20. Dang, L. C. *et al*. Spontaneous Eye Blink Rate (EBR) Is Uncorrelated with Dopamine D2 Receptor Availability and Unmodulated by Dopamine Agonism in Healthy Adults. eNeuro 4. *ENEURO*. **0211–17**, 2017–11 (2017).
21. Van Slooten, J. C., Jahfari, S., Knapen, T. & Theeuwes, J. Individual differences in eye blink rate predict both transient and tonic pupil responses during reversal learning. *PLOS ONE* **12**, e0185665–20 (2017).
22. Jahfari, S. *et al*. Cross-Task Contributions of Frontobasal Ganglia Circuitry in Response Inhibition and Conflict-Induced Slowing. *Cerebral Cortex* **4**, 1–15 (2018).
23. Hamid, A. A. *et al*. Mesolimbic dopamine signals the value of work. *Nature Neuroscience* **19**, 117–126 (2015).
24. Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L. & Platt, M. L. A Primer on Foraging and the Explore-Exploit Trade-Off for. *Psychiatry Research.* **42**, 1931–1939 (2017).
25. Berke, J. D. What does dopamine mean? *Nature Neuroscience* **21**, 787–793 (2018).
26. Jahfari, S. & Theeuwes, J. Sensitivity to value-driven attention is predicted by how we learn from value. *Psychonomic Bulletin Review* **24**, 408–415 (2016).
27. Van Slooten, J. C., Jahfari, S., Knapen, T. & Theeuwes, J. How pupil responses track value-based decision-making during and after reinforcement learning. *PLOS Comput Biol* **14**, e1006632–25 (2018).
28. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nature Neuroscience* **10**, 1214–1221 (2007).
29. Shen, W., Flajolet, M., Greengard, P. & Surmeier, D. J. Dichotomous Dopaminergic Control of Striatal Synaptic Plasticity. *Science* **321**, 848–851 (2008).
30. 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).
31. Kahnt, T. *et al*. Dorsal Striatalmidbrain Connectivity in Humans Predicts How Reinforcements Are Used to Guide Decisions. *Journal of Cognitive Neuroscience* **21**, 1332–1345 (2009).
32. McCoy, B., Jahfari, S., Engels, G., Knapen, T. & Theeuwes, J. Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. *Brain*, **142**, 3605–3620 (2019).
33. 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).
34. 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).
35. 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).
36. 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).
37. Morris, T. L. & Miller, J. C. Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology* **42**, 343–360 (1996).
38. Häkkänen, H., Summala, H., Partinen, M., Tiihonen, M. & Silvo, J. Blink Duration as an Indicator of Driver Sleepiness in Professional Bus Drivers. *Sleep* **22**, 798–802 (1999).
39. 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).
40. 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).
41. Naurois, C. J., de, Bourdin, C., Stratulat, A., Diaz, E. & Vercher, J.-L. Detection and prediction of driver drowsiness using artificial neural network models. *Accident Analysis and Prevention* **126**, 95–104 (2019).
42. Wilson, R. C. & Collins, A. G. E. Ten simple rules for the computational modeling of behavioral data. *psyRxiv* 1–35, https://doi.org/10.31234/osf.io/46mbn (2019).
43. 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. *Decision* 1–49 (2017).
44. Jeffreys, H. *Theory of Probability*. (Oxford: Oxford University Press, 1961).
45. Epskamp, S. & Fried, E. I. A Tutorial on Regularized Partial Correlation Networks. *Psychological Methods*, **23**, 617–634 (2018).
46. Rac-Lubashevsky, R., Slagter, H. A. & Kessler, Y. Tracking Real-Time Changes in Working Memory Updating and Gating with the Event-Based Eye-Blink Rate. *Scientific Reports* **7**, 343–9 (2017).
47. Dreisbach, G. *et al*. Dopamine and Cognitive Control: The Influence of Spontaneous Eyeblink Rate and Dopamine Gene Polymorphisms on Perseveration and Distractibility. *Behavioral Neuroscience* **119**, 483–490 (2005).
48. Müller, J. *et al*. Dopamine and cognitive control: The influence of spontaneous eyeblink rate, DRD4 exon III polymorphism and gender on flexibility in set-shifting. *Brain Research* **1131**, 155–162 (2007).
49. Tharp, I. J. & Pickering, A. D. Individual differences in cognitive-flexibility: The influence of spontaneous eyeblink rate, trait psychoticism and working memory on attentional set-shifting. *Brain and cognition* **75**, 119–125 (2011).
50. Frank, M. J. Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience* **17**, 51–72 (2005).
51. Doll, B. B. & Frank, M. J. The basal ganglia in reward and decision making: computational models and empirical studies. In *Handbook of reward and decision making,* 399–425 (Elsevier Inc. 2009).
52. Cohen, M. X. & Frank, M. J. Neurocomputational models of basal ganglia function in learning, memory and choice. *Behavioural Brain Research* **199**, 141–156 (2009).
53. Beeler, J. A. Putting desire on a budget: dopamine and energy expenditure, reconciling reward and resources. Frontiers in Integrative. *Neuroscience* **6**, 1–22 (2012).
54. Salamone, J. D. & Correa, M. The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron* **76**, 470–485 (2012).
55. 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).
56. Cagniard, B. *et al*. Dopamine Scales Performance in the Absence of New Learning. *Neuron* **51**, 541–547 (2006).

57. Beeler, J. A., Daw, N., Frazier, C. R. M. & Zhuang, X. Tonic Dopamine Modulates Exploitation of Reward Learning. Frontiers in. *Behavioral Neuroscience* **4**, 1–14 (2010).
58. Salamone, J. D., Wisniecki, A., Carlson, B. & Correa, M. Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. *Neuroscience* **105**, 863–870 (2001).
59. Humphries, M. D., Khamassi, M. & Gurney, K. Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Frontiers in Neuroscience* **6**, (2012).
60. 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).
61. Cinotti, F. *et al*. Dopamine blockade impairs the exploration-exploitation trade-off in rats. *Scientific Reports* **9**, 1–14 (2019).
62. Chakroun, K., Mathar, D., Wiehler, A., Ganzer, F. & Peters, J. Dopaminergic modulation of the exploration/exploitation trade-off in human decision-making. *bioRxiv* 1–55, https://doi.org/10.1101/706176 (2019).
63. Grogan, J. P. *et al*. Effects of dopamine on reinforcement learning and consolidation in Parkinsons disease. *eLife* **6**, 14491 (2017).
64. Shiner, T. *et al*. Dopamine and performance in a reinforcement learning task: evidence from Parkinsons disease. *Brain* **135**, 1871–1883 (2012).
65. Smittenaar, P. *et al*. Decomposing effects of dopaminergic medication in Parkinsons disease on probabilistic action selection: learning or performance? *European Journal of Neuroscience* **35**, 1144–1151 (2012).
66. McCoy, B., Jahfari, S., Knapen, T. & Theeuwes, J. Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. *Brain* 1–68 (2019).
67. Jiang, X., Tien, G., Huang, D., Zheng, B. & Atkins, M. S. Capturing and evaluating blinks from video-based eyetrackers. *Behavior Research Methods* **45**, 656–663 (2012).
68. Barbato, G. *et al*. Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research* **93**, 145–151 (2000).
69. Frank, M. J., Seeberger, L. C. & O'reilly, R. C. By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940–1943 (2004).
70. Watkins, C. J. C. H. & Dayan, P. Technical Note: Q-Learning. *Machine Learning* **8**, 279–292 (1992).
71. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**, 1413–1432 (2016).
72. Fontanesi, L., Gluth, S., Spektor, M. S. & Rieskamp, J. A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin Review* 1–23, https://doi.org/10.3758/s13423-018-1554-2 (2019).
73. Lee, M. D. How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology* **55**, 1–7 (2011).
74. Gelman, A. *et al*. *Bayesian Data Analysis*., https://doi.org/10.1201/b16018 (Chapman; Hall/CRC, 2013).
75. Scheibehenne, B. & Pachur, T. Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin Review* **22**, 391–407 (2014).
76. Lee, M. D., Lodewyckx, T. & Wagenmakers, E.-J. Three Bayesian Analyses of Memory Deficits in Patients with Dissociative Identity Disorder. In *Cognitive modeling in perception and memory*. 189–200 (2014).

Acknowledgements

We thank Lisa Roodermond and Lynn van den Berg for their assistance in the data collection of this study. We thank Tomas Knapen for discussion and comments on an earlier draft of the paper. This research is funded by the ERC advanced grant [ERC-2012-AdG-323413] to J.T.

Author contributions

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

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-53805-y.

Correspondence and requests for materials should be addressed to J.C.V.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.