


## Serotonin modulates asymmetric learning from reward and punishment in healthy human volunteers

Jochen Michely <sup>1,2,3,4✉</sup>, Eran Eldar <sup>5</sup>, Alon Erdman <sup>5</sup>, Ingrid M. Martin<sup>4,6</sup> & Raymond J. Dolan <sup>3,4</sup>

Instrumental learning is driven by a history of outcome success and failure. Here, we examined the impact of serotonin on learning from positive and negative outcomes. Healthy human volunteers were assessed twice, once after acute (single-dose), and once after prolonged (week-long) daily administration of the SSRI citalopram or placebo. Using computational modelling, we show that prolonged boosting of serotonin enhances learning from punishment and reduces learning from reward. This valence-dependent learning asymmetry increases subjects' tendency to avoid actions as a function of cumulative failure without leading to detrimental, or advantageous, outcomes. By contrast, no significant modulation of learning was observed following acute SSRI administration. However, differences between the effects of acute and prolonged administration were not significant. Overall, these findings may help explain how serotonergic agents impact on mood disorders.

<sup>1</sup>Department of Psychiatry and Neurosciences, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany. <sup>2</sup>Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Charité Clinician Scientist Program, Berlin, Germany. <sup>3</sup>Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK. <sup>4</sup>Wellcome Centre for Human Neuroimaging, University College London, London, UK. <sup>5</sup>Psychology and Cognitive Sciences Departments, Hebrew University of Jerusalem, Jerusalem, Israel. <sup>6</sup>Institute of Cognitive Neuroscience, University College London, London, UK. ✉email: [jochen.michely@charite.de](mailto:jochen.michely@charite.de)

To maximize reward, and minimize punishment, agents need to learn from a history of past success and failure<sup>1</sup>. Evidence suggests that, rather than being represented on a continuous scale, reward and punishment may represent distinct categorical events<sup>2,3</sup>. This is corroborated by findings that exposure to reward and punishment engage distinct brain networks<sup>4–7</sup>. This triggers distinct types of behaviour, such as approach and invigoration for reward, and avoidance and inhibition for punishment<sup>8,9</sup>.

Numerous studies report an asymmetric impact of positive and negative outcomes during human instrumental learning<sup>10–12</sup>. Interestingly, such valence-dependent learning asymmetries are characterised by remarkable flexibility, for example an adjustment to environmental volatility<sup>13</sup> or contextual information<sup>14</sup>. Moreover, learning asymmetries are linked to interindividual variability in brain structure and function<sup>7,15</sup>, and are thought to play a role in the emergence of mood disorders, often characterised in terms of aberrant processing of reward and punishment<sup>16,17</sup>.

Previous research shows that the neuromodulators dopamine and serotonin play a key role in modulating reward and punishment learning. Whilst there is ample evidence for a role of dopamine in learning from reward<sup>18–21</sup>, the evidence in relation to serotonin is less clear. Some studies indicate a specific role in punishment learning<sup>22–26</sup>, while others report that serotonin impacts learning from both reward and punishment<sup>27,28</sup>.

A key challenge studying the human serotonergic system is the manipulation of central serotonin levels. In previous studies, serotonin has often been modulated via a dietary depletion of tryptophan, a precursor of serotonin<sup>29</sup>. However, conclusive evidence supporting the effectiveness and specificity of this dietary manipulation in humans is lacking<sup>30</sup>, but cf.<sup>31</sup>.

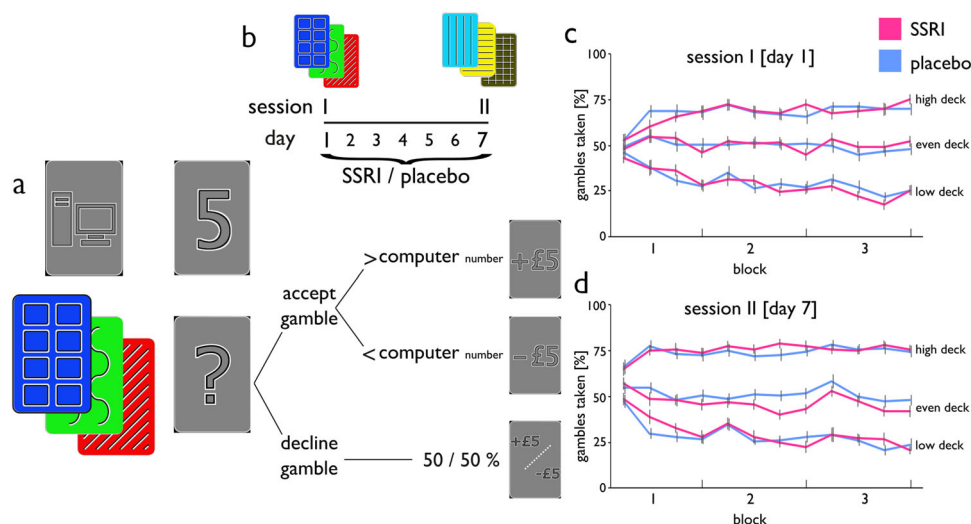
Arguably, a more specific method involves a pharmacological enhancement of central serotonin via the use of selective serotonin reuptake inhibitors (SSRIs). To date, SSRI studies of

learning asymmetries have been mainly limited to single dose administration or assessment of clinical populations, which makes interpretation problematic<sup>26,27,32</sup>. Human and non-human animal data suggest a pharmacological modulation of serotonin can impact brain function at different timescales, with distinct effects of single dose, one-day intervention, and prolonged, repeated administration over several days<sup>33–37</sup>. This accords with the fact that SSRIs reach steady-state peak plasma levels only after a prolonged treatment spanning multiple days<sup>38</sup>. This renders it likely that prolonged and repeated administration of serotonergic drugs is necessary to impact on behaviour, such as inducing a substantial modulation of learning processes<sup>28,39</sup>.

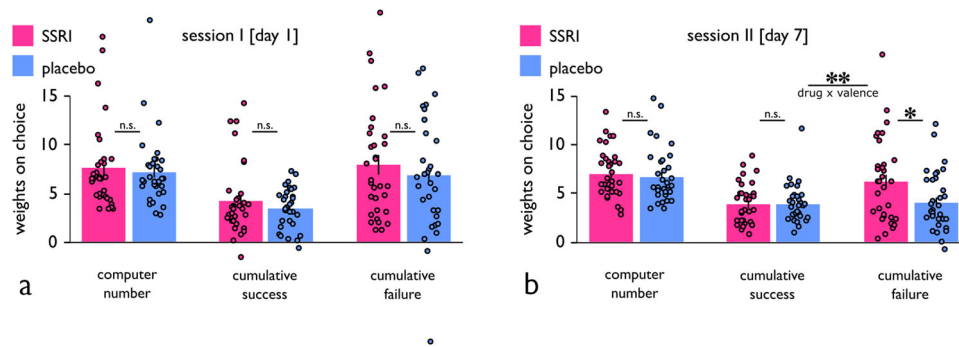
Based on these considerations, we examined the impact of an extended exposure to a serotonergic treatment on human reinforcement learning. Healthy human volunteers were assessed twice, once after single dose, and once after repeated daily administration of either 20 mg of the SSRI citalopram, or placebo, across seven consecutive days. Subjects performed a task specifically tailored to study an asymmetry in learning from reward and punishment<sup>7</sup>. We used computational modelling to examine fine-grained characteristics of learning over time. We show prolonged SSRI treatment exerts an asymmetric impact on reinforcement learning, reducing learning from reward and enhancing learning from punishment. We discuss the implications of these findings with respect to potential mode of action of serotonergic treatment in the context of mood disorders.

## Results

We administered sixty-six healthy volunteers either a daily oral dose of the SSRI citalopram (20 mg) or placebo, across seven consecutive days. Subjects performed two experimental sessions, once after administration of a single dose on day 1 and once, after exposure to repeated daily administration, on day 7 (Fig. 1b).



**Fig. 1 Experimental task, pharmacological procedure, and learning performance.** **a** Experimental design: On each trial, participants were presented with one of three possible decks and a number between 1 and 9 drawn by the computer. If participants decided to gamble, they won £5 if the number they drew was higher than the computer drawn number, and lost £5 if the number was lower. Participants were only informed whether they won or lost the gamble, not which number they drew. Participants had to learn by trial and error how likely gambles were to succeed with each of the three decks. One deck contained a uniform distribution of numbers between 1 and 9 (even deck), one deck contained more 1's (low deck), making gambles 30% less likely to succeed, and one deck contained more 9's (high deck), making gambles 30% more likely to succeed. Opting to decline the gamble resulted in a 50% probability of win/loss, regardless of which number was drawn by the computer. **b** Pharmacological procedure: Subjects were randomly allocated to take a daily dose of 20 mg citalopram or placebo for seven days and participated in two sessions: session I took place on day 1 after single administration, session II took place on day 7, at a time when citalopram reaches steady-state plasma levels. Subjects played an identical game on both sessions, but with two independent sets of three decks, where colour order and colour-associated win probability randomly varied across participants. **c, d** Learning performance: Gambles taken with each deck as a function of time. Percentages were computed separately for each set of 15 contiguous trials (4 sets/60 trials per block), for session I (**c**), and session II (**d**), respectively. Error bars indicate SEM.



**Fig. 2 Results of trial-by-trial logistic regression model.** Fitting a logistic regression model to subjects' decisions showed that participants gambled more against lower computer numbers, with no drug differences on session I (**a**), and session II (**b**), respectively. Additionally, subjects, over the course of a session, gambled more with increasing success with each deck, and gambled less with increasing failure with each deck. On session I, impact of cumulative success and failure was unaffected by treatment (**a**). On session II, however, SSRIs enhanced the impact of failure but not wins (**b**), indicating an asymmetric drug effect on reward and punishment. \*\* $p < 0.01$ , \* $p < 0.05$ , n.s. = not significant ( $p > 0.05$ ). Error bars indicate SEM.

During each session, subjects performed a modified version of a gambling card game (Fig. 1a;<sup>7</sup>), where the goal was to maximize monetary wins and minimize monetary losses.

In brief, on each trial participants were presented with a number between 1 and 9 as drawn by a computer. Subjects could gamble that the number they were about to draw would be higher than this computer drawn number. Critically, participants played with one of three possible decks on each trial, where the decks differed in how likely gambles were to succeed. One deck contained a uniform distribution of numbers between 1 and 9 (even deck), one deck contained more 1's (low deck, gambles 30% less likely to be successful), and one deck contained more 9's (high deck, gambles 30% more likely to be successful).

Subjects were informed that an unsuccessful gamble would result in a loss ( $-\pounds 5$ ), and a successful gamble would result in a win ( $+\pounds 5$ ). Subjects learnt through trial and error about each of the decks' success likelihood. Alternatively, subjects could decline a gamble and instead opt for a fixed 50% known probability of winning or losing, respectively. After a decision to decline a gamble, the outcome was not shown to participants.

On a second session, the game was identical, and the only difference being that subjects played with three novel decks, indicated by different colours, where colour order and colour-associated win probability was randomly varied across participants (Fig. 1b). Subjects had to learn about these decks anew as they were unrelated to the ones from the first session.

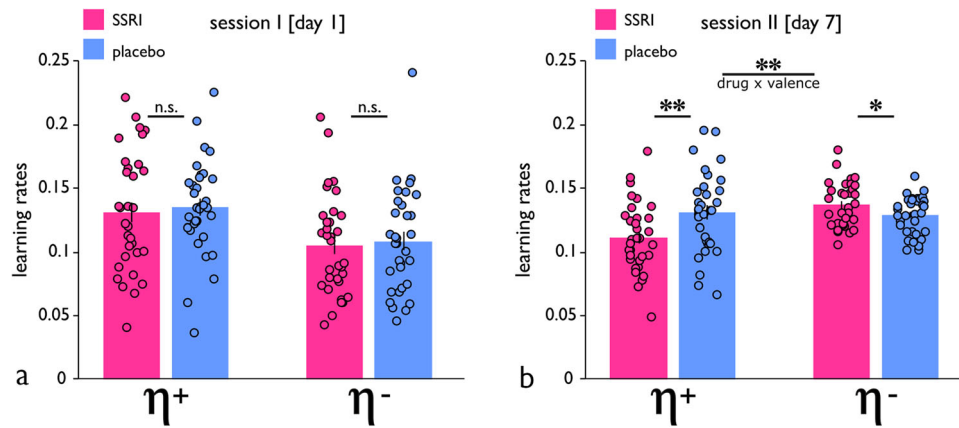
The experiment was designed such that the computer numbers changed over time to ensure subjects gambled on approximately 50% of trials across all decks (cf. Methods). Indeed, this adaptation worked, and overall proportion of accepted gambles did not differ between drug groups (session I: SSRI 49.0%, placebo 49.3%,  $t_{65} = 0.1$ ,  $p = 0.846$ ; session 2: SSRI 50.6%, placebo 51.2%,  $t_{65} = 0.4$ ,  $p = 0.641$ ). Thus, evidence of learning manifested in how the rate of accepted gambles differed between decks, and in an observation that this difference grew over the course of the experiment, i.e., from 1st to 3rd block (Fig. 1c/d; session I: low vs. even,  $t_{65} = 3.2$ ,  $p = 0.0016$ ; low vs. high,  $t_{65} = 6.1$ ,  $p = 6.0e-8$ ; even vs. high,  $t_{65} = 3.0$ ,  $p = 0.003$ ; session II: low vs. even,  $t_{65} = 2.9$ ,  $p = 0.003$ ; low vs. high,  $t_{65} = 6.1$ ,  $p = 6.2e-8$ ; even vs. high,  $t_{65} = 2.7$ ,  $p = 0.007$ ). There was no significant difference between the groups in this respect (session I: low vs. even, SSRI vs. placebo:  $t_{64} = 1.4$ ,  $p = 0.159$ ; low vs. high, SSRI vs. placebo:  $t_{64} = 1.0$ ,  $p = 0.291$ ; even vs. high, SSRI vs. placebo:  $t_{64} = -0.2$ ,  $p = 0.783$ ; session II: low vs. even, SSRI vs. placebo:  $t_{64} = 0.06$ ,  $p = 0.949$ ; low vs. high, SSRI vs. placebo:  $t_{64} = 0.6$ ,  $p = 0.491$ ; even vs. high, SSRI vs. placebo:  $t_{64} = 0.6$ ,  $p = 0.549$ ). Overall, this demonstrates that subjects learned to dissociate decks, as their

willingness to gamble differed depending on each deck's win likelihood as a function of time, and this effect was not modulated by the drug.

Next, we used a trial-by-trial logistic regression approach (cf. Methods) to assess whether subjects' decisions to gamble were dependent upon the computer number and previous receipt of positive, or negative, outcomes over time. First, we found that subjects, over both sessions, gambled more against lower computer numbers (session I:  $t_{64} = 14.7$ ,  $p = 3.0e-22$ ; session II:  $t_{65} = 20.9$ ,  $p = 1.3e-30$ ; Fig. 2), with no evidence for a difference between drug groups (drug:  $F_{1,63} = 0.4$ ,  $p = 0.505$ ; drug  $\times$  session:  $F_{1,63} = 0.06$ ,  $p = 0.805$ ).

Second, participants, over the course of each session, gambled more with decks with which they had experienced more success (session I:  $t_{64} = 10.3$ ,  $p = 3.0e-15$ ; session II:  $t_{65} = 15.7$ ,  $p = 8.2e-24$ ) and less failure (session I:  $t_{64} = 10.2$ ,  $p = 4.3e-15$ ; session II:  $t_{65} = 11.0$ ,  $p = 1.6e-16$ ). This result indicates subjects successfully learned about the decks from the outcomes of their gambles. When assessing data across both sessions, the pharmacological effect on gambling preferences as a function of outcome type was not statistically significant (drug:  $F_{1,63} = 1.7$ ,  $p = 0.194$ , drug  $\times$  valence:  $F_{1,63} = 2.6$ ,  $p = 0.108$ , drug  $\times$  session  $\times$  valence:  $F_{1,63} = 2.4$ ,  $p = 0.124$ ). However, analysing both sessions separately, effects were similar across drug groups for cumulative success and failure on session I (drug  $\times$  valence:  $F_{1,63} = 0.3$ ,  $p = 0.844$ ; drug:  $F_{1,63} = 0.9$ ,  $p = 0.345$ ; Fig. 2a), whereas on session II we found evidence for an asymmetric impact of success and failure outcomes, as a function of treatment (drug  $\times$  valence,  $F_{1,64} = 10.5$ ,  $p = 0.0018$ ; drug:  $F_{1,64} = 2.4$ ,  $p = 0.126$ ; Fig. 2b), attributable to an enhanced impact of failure ( $t_{64} = 2.3$ ,  $p = 0.024$ ) but not of success ( $t_{64} = 0.1$ ,  $p = 0.892$ ), in SSRI treated as compared to placebo subjects. This differential pattern suggests that SSRI treatment increased an impact of negative outcomes, enhancing a gamble avoidance tendency in response to failure.

Next, we used computational modelling (cf. Methods) to assess the precise learning mechanism underlying the asymmetric effects of success and failure evident in the regression analysis. Replicating results from an earlier study using an identical cognitive task<sup>7</sup>, model comparison showed task behaviour was best explained by a model that accounted for an asymmetry in learning from the two outcome types. Specifically, the best-fitting model included two different learning rates, one for wins ( $\eta^+$ ), and one for losses ( $\eta^-$ ), where these determine the degree to which an outcome type impacts on subsequent expectations (model 6: 'adjusted & asymmetric Q-learning'; cf. Supplementary Table 1 for iBIC scores). These expectations, in combination with the numbers drawn by the computer, shape whether gambles are



**Fig. 3 Learning asymmetry and its serotonergic modulation.** **a** On session I, computational modelling showed learning from reward ( $\eta^+$ ) and learning from punishment ( $\eta^-$ ) was unaffected by drug treatment. **b** On session II, SSRI treatment had a significant effect on learning asymmetries, such that it reduced learning from reward ( $\eta^+$ ), and enhanced learning from punishment ( $\eta^-$ ). Note that data simulation showed that an ‘optimal’ learning rate for the task, maximizing net reward gained, was in the range of  $\approx 0.3$ – $0.6$  (Supplementary Fig. 2). \*\* $p < 0.01$ , \* $p < 0.05$ , n.s. = not significant ( $p > 0.05$ ). Error bars indicate SEM.

likely to be taken or declined. The predictive accuracy of the model (absolute fit), i.e., the proportion of subjects’ choices to which the model gives a likelihood greater than 50% (percent correct), was 87.71% for session I, and 87.92% for session II (Supplementary Fig. 1).

When assessing computational parameters across data from both sessions, we found a significant asymmetric effect of SSRIs on learning rates (drug  $\times$  valence:  $F_{1,62} = 4.1$ ,  $p = 0.046$ ; drug:  $F_{1,62} = 0.8$ ,  $p = 0.365$ ), but no significant three-way interaction (drug  $\times$  valence  $\times$  session:  $F_{1,62} = 1.0$ ,  $p = 0.305$ , controlling for an overall gambling bias, Supplementary Fig. 3). Follow-up tests revealed that, on session I, there was no evidence for computational parameters governing the rate of learning from reward and punishment being different between treatment groups (drug  $\times$  valence:  $F_{1,64} = 0.007$ ,  $p = 0.933$ ; drug:  $F_{1,64} = 0.3$ ,  $p = 0.553$ ; Fig. 3a). However, by session II, a significant serotonergic impact on learning asymmetry was evident (drug  $\times$  valence:  $F_{1,64} = 8.2$ ,  $p = 0.006$ ; drug:  $F_{1,64} = 3.2$ ,  $p = 0.075$ ; Fig. 3b), such that in SSRI, as compared to placebo subjects, learning from reward was reduced ( $t_{64} = 2.7$ ,  $p = 0.008$ ) while learning from punishment was enhanced ( $t_{64} = 2.0$ ,  $p = 0.041$ ).

Overall, these results indicate that a prolonged regimen of SSRI treatment resulted in a modulation of learning asymmetries. Importantly, there were no between-group differences for the remaining model parameters (Supplementary Fig. 3). With regards to the impact of a single SSRI dose, on the one hand there was no significant impact following a single dose, while on the other the impact following a single dose did not significantly differ from the impact following prolonged SSRI treatment.

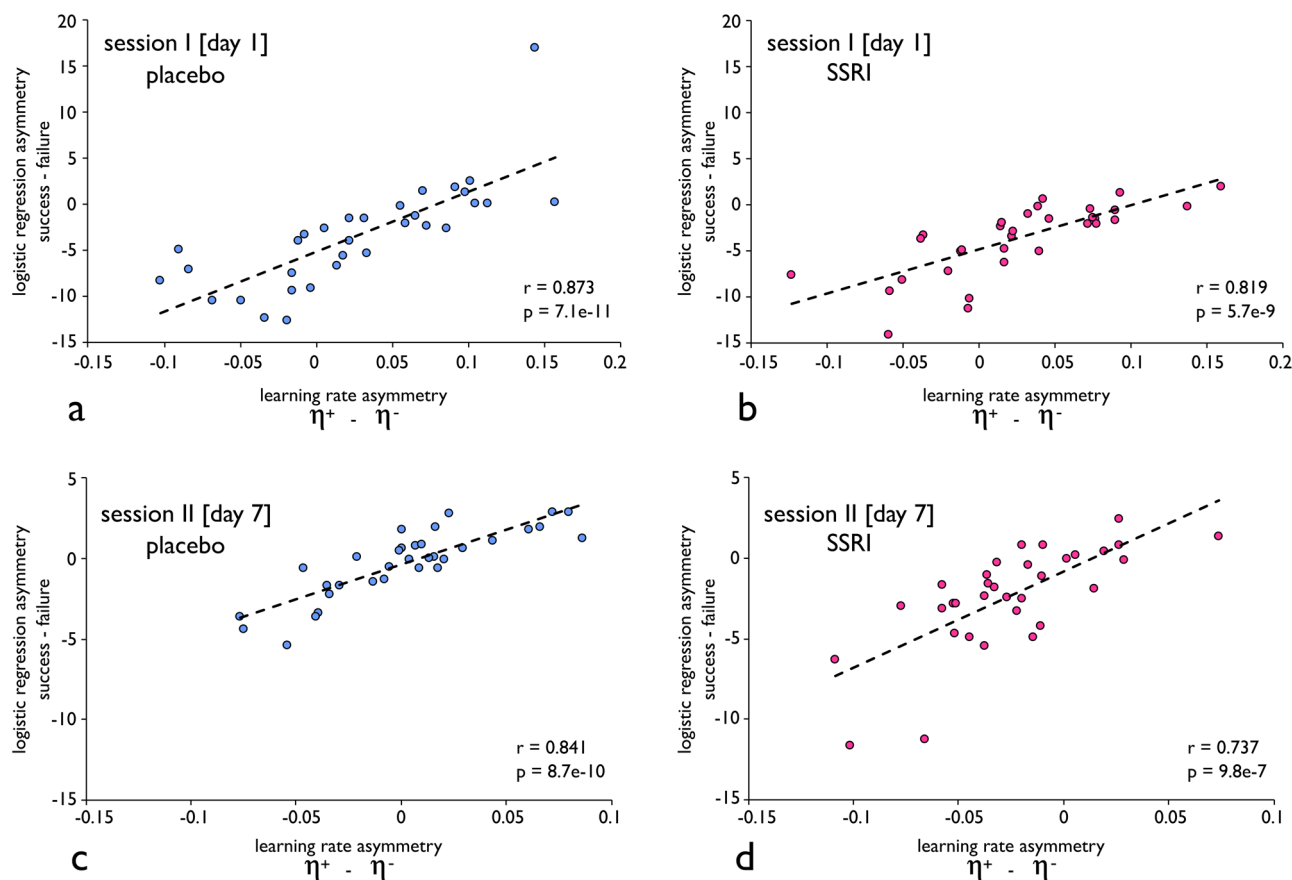
Additionally, an asymmetric effect of cumulative success and failure on gambling, as derived from the logistic regression, correlated significantly with an asymmetry in learning, as derived from the computational reinforcement learning model, in both sessions for both drug groups (session I, placebo:  $r = 0.873$ ,  $p = 7.1e-11$ , session I, SSRI:  $r = 0.819$ ,  $p = 5.7e-9$ ; session II, placebo:  $r = 0.841$ ,  $p = 8.7e-10$ ; session II, SSRI:  $r = 0.737$ ,  $p = 9.8e-7$ ; Fig. 4).

The benefit of having both analyses is that the model-based analysis is more sensitive, albeit at a cost of greater flexibility in fitting the data. Specifically, reinforcement learning modelling can mimic our regression analysis by fitting the data with very low learning rates, thus weighting outcomes almost equally. However, by fitting the data with higher learning rates, it can also place substantially greater weight on recent outcomes. We additionally illustrate the correspondence between these two measures

(regression and reinforcement learning modelling) in simulations with a wide range of parameter settings. Briefly, we simulated artificial data from five models, in which we randomly varied positive and negative learning rates independently across agents. Next, we ran logistic regression analyses on the artificial data and computed correlations between an asymmetric effect of cumulative success and failure on gambling (regression) and an asymmetry in learning (computational model). Here, we found a highly significant relationship across all simulated data sets ( $r$  ranging between 0.82–0.87, all  $p < 2.6e-17$ ), providing further evidence for the relationship between these two measures. Overall, these analyses jointly indicate asymmetric learning from positive and negative outcomes related to an altered gambling preference and was influenced by prolonged serotonergic intervention.

Note in addition we tested a model with differing sensitivity to outcome valence (positive and negative, respectively), and a model that modulates both a sensitivity to distinct outcomes and learns differently from these distinct outcomes. Across both sessions, the models provided a worse fit than a model that learns asymmetrically from distinct outcomes (Supplementary Table 1). Although the latter model was clearly not the best-fitting model, we used it for a joint test of drug effects on both outcome sensitivity and learning parameters from session II. Here we found that a sensitivity to outcomes did not differ between drug groups ( $t_{64} = 0.5$ ,  $p = 0.582$ ), but that asymmetric learning was modulated by SSRIs (drug  $\times$  valence:  $F_{1,64} = 10.4$ ,  $p = 0.0019$ ), with reduced learning from reward ( $t_{64} = 2.9$ ,  $p = 0.004$ ), and an increased learning from punishment ( $t_{64} = 2.1$ ,  $p = 0.032$ ) following serotonergic intervention, mirroring the drug effects on parameters of our winning model.

Note that, across all subjects and sessions, we found no significant difference between positive and negative learning rates ( $F_{1,65} = 2.1$ ,  $p = 0.151$ ). However, in placebo subjects alone, we found a statistical trend for greater learning from positive as compared to negative outcomes ( $F_{1,32} = 3.7$ ,  $p = 0.063$ ). This is in line with recent work showing a learning asymmetry towards greater updating from positive information healthy individuals, without pharmacological intervention<sup>15</sup>. We found no difference between drug groups in net reward gained, a key measure for task performance (session I:  $t_{64} = 1.2$ ,  $p = 0.201$ ; session II:  $t_{64} = 0.4$ ,  $p = 0.650$ ). Thus, we found no evidence that changes in asymmetric learning were detrimental, or advantageous, for task performance.



**Fig. 4 Asymmetric effects of reward and punishment.** An asymmetric effect of cumulative success and failure on gambling, as derived from the logistic regression, significantly correlated with an asymmetry in learning, as derived from our computational model, in both sessions for both drug groups.

**a** Session I, placebo. **b** Session I, SSRI. **c** Session II, placebo. **d** Session II, SSRI.

We performed several analyses to assess the validity of our computational modelling approach. First, we generated simulated data based upon model parameters derived from fitting to real data. This ‘posterior predictive check’ confirmed that the model captured core features of the real data (Supplementary Fig. 4). Additionally, we simulated sets of choices from artificial agents with specific sets of parameters (‘ground truth’) and then fitted models to those choices to recover the values of the parameters (‘recovered parameters’). To ensure the results of the parameter recovery test were applicable to the analysis of the real data, we selected the ground truth parameters such that they covered the empirical range. This analysis revealed that individual parameter estimates could be accurately recovered (Supplementary Figs. 5/6). Lastly, we validated our model comparison procedure by generating simulated data using each model and applying our model comparison procedure to identify the model that generated each dataset (Supplementary Fig. 7).

## Discussion

Here, we show that boosting central serotonin by means of week-long SSRI administration enhanced learning from punishment and reduces learning from reward. This SSRI-induced learning asymmetry increased subjects’ tendency to avoid actions as a function of cumulative failure.

Serotonin is an evolutionary conserved neurotransmitter though its precise effects on cognition has evaded a definite mechanistic understanding<sup>40,41</sup>. One influential proposal is that serotonin plays a specific role in processing aversive outcomes<sup>42</sup>. Indeed, several studies in humans show that serotonin is involved

in punishment learning<sup>22–26</sup>, but other studies suggest that it impacts learning from both positive and negative outcomes<sup>27,28</sup>.

In our study, we replicate findings from a previous study that used an identical learning task<sup>7</sup>, showing behaviour is best explained by an asymmetry in reward and punishment learning. A strength of our task is that learning from reward and punishment are each assessed via their naturally associated go (i.e., accept gamble) or no-go (i.e., reject gamble) Pavlovian responses<sup>9</sup>. Additionally, reward and punishment are administered within the same block, in an interleaved manner, thereby competing for a subject’s learning resources<sup>43</sup>. Unfortunately, in this study, we were not in a position to acquire neural data. In light of previous studies on interindividual variability in human learning asymmetries, it is tempting to speculate that serotonergic agents may act preferentially in the striatum and prefrontal cortex to alter the relative degree of impact from positive and negative outcomes<sup>7,15,44</sup>.

The effects of a serotonergic manipulation we highlight require a temporally extended treatment in order to evolve. This accords with human and non-human animal studies showing that only a prolonged modulation of serotonin induces a substantial impact<sup>33,36,45,46</sup>, particularly with respect to learning<sup>28,35</sup>. The fact that changes in learning emerge after an extended intervention may reflect two processes, or a synergism of both. First, citalopram reaches steady-state plasma levels after seven days<sup>38</sup>, and a single dose administration is unlikely to suffice for induction of a substantial modulation of learning. Second, plasticity that may underlie this modulation, such as neurogenesis, synaptogenesis, or changes in BDNF levels, require days or weeks to emerge<sup>47</sup>.

A learning asymmetry, involving a greater impact of losses than wins, can lead to increased avoidance relative to approach behaviour. This can result in an aversion to risk-taking and action over time, and potentially maladaptive risk-avoidant behaviour<sup>48</sup>. However, studies specifically assessing the impact of serotonin on human risk-taking have, thus far, proven inconclusive<sup>29,49–51</sup>. Notably, in these studies, decision variables are typically not learned by trial and error but are instead presented to subjects explicitly, which contrasts with the learning design utilised in the current study. Thus, the relationship between a serotonergic effect on asymmetric learning and the development of risk tendencies remains a question for pursuit in future studies.

Aberrant processing of reward and punishment is assumed to play a role in the emergence of mood disorders<sup>16,17</sup>. Although SSRIs constitute a first-line pharmacological intervention in mood disorders<sup>52</sup>, the cognitive and computational mechanisms that underpin treatment effectiveness remain an unresolved issue<sup>53</sup>. We suggest the asymmetric effects we highlight can help explain the clinical impact of SSRIs. Specifically, recent lab-based<sup>54–57</sup> and real-world studies<sup>58,59</sup> show that outcome prediction errors, or positive and negative surprise, strongly impact on subjects' emotional state. Here, an agent's mood depends not only on how well things are going in general, but whether things are better, or worse, than expected. Our results indicate that a serotonergic intervention can, in principle, influence the affective impact of reinforcement, by lowering positive expectations through slowing reward learning, thereby giving rise to more positive surprise, and minimizing the impact of aversive outcomes by enhancing negative expectations as a function of increased punishment learning. Overall, a consequence is an increase in positive as compared to negative surprise, an effect that may contribute to a gradual emergence of better mood<sup>60</sup>.

A related line of work shows that healthy individuals learn more from positive, relative to negative, information leading to an 'optimism bias'<sup>61,62</sup>, where the latter is lacking in depressed patients<sup>63,64</sup>. This might seem to suggest that an increased learning from positive information acts to protect against depression. It is worth highlighting, however, that these studies typically assay updating expectations about oneself given information about an average person. This type of assay is important, but it involves an additional critical factor—the degree to which subjects accept that information about the average applies to themselves. There are a range of reasons to discount the relevance of average statistics, and such discounting can be applied asymmetrically to positive and negative information. We propose this factor, rather than differences in learning per se, might explain depression-related individual differences in optimism bias. In contrast, our approach strives to assess a more basic process of learning from reward and punishment. For this we designed our study such that the feedback from which subjects learn is unequivocal, as is typically applied in the basic reinforcement learning literature<sup>7,10,11</sup>.

Overall, we consider that results, alongside its putative impact on changes in mood, do not contradict findings on optimism bias in depression. On the contrary, both processes can be expected to contribute to an emergence of better mood<sup>60</sup>. Here, a limitation of our study is its restriction to non-depressed healthy individuals. Moreover, week-long SSRI treatment does not typically result in a meaningful mood improvement in a clinically depressed population<sup>65</sup>. Furthermore, our task did not contain a concurrent mood measurement. Ultimately, testing both self-referential and basic reinforcement learning, alongside tracking of subjective changes in mood in clinical cohorts, will be an important next step for examining a temporal evolution of changes in learning and the emergence of clinical effects over time.

Although our data suggests an emergence of serotonergic effects after a temporally extended intervention, it is of note that a three-way interaction (drug  $\times$  session  $\times$  valence) was not significant. Thus, we tentatively conclude that prolonged treatment induced a learning asymmetry, but the interpretation of this needs to be tempered by the fact that there was no difference between prolonged (week-long, on day 7) as compared to acute (single-dose, day 1) treatment. To unravel the precise trajectory of any such effect, future studies should ideally include a pre-drug testing session as well as multiple sessions over several weeks of treatment.

In summary, we show that week-long SSRI treatment reduces reward and enhances punishment learning. This learning asymmetry can, in theory, result in lowered positive and enhanced negative expectations, and consequentially, to more rewarding, and less disappointing experience. We suggest this modulation of computations that guide reinforcement learning may contribute to a known serotonergic impact on mood.

## Methods

**Subjects.** Sixty-six healthy volunteers (mean age:  $24.7 \pm 3.9$ ; range 20–38 years; 40 females; Supplementary Table 2) participated in this double-blind, placebo-controlled study. All subjects underwent an electrocardiogram to exclude QT interval prolongation and a thorough medical screening interview to exclude any neurological or psychiatric disorder, any other medical condition, or medication intake. Subjects were reimbursed for their time. Additionally, subjects were informed that, at the end of the experiment, one trial was randomly selected, and the outcome of that trial was added to the overall payment. Thus, performance was incentivised as choosing good gambles resulted in a higher probability of earning additional monetary reward. Data from different tasks of the same participants are reported elsewhere<sup>66,67</sup>. The experimental protocol was approved by the University College London (UCL) local research ethics committee, with informed consent obtained from all participants.

**Pharmacological procedure.** Participants were randomly allocated to receive a daily oral dose of the SSRI citalopram (20 mg) or placebo, over a period of seven consecutive days. All subjects performed two laboratory testing sessions. Session I was on day 1 of treatment, 3 h after single dose administration, as citalopram reaches its highest plasma levels after this interval<sup>68</sup>. On the following days, subjects were asked to take their daily medication dose at a similar time of day, either at home or at the study location. Session II was on day 7 of treatment, a time when citalopram is known to reach steady-state plasma levels<sup>38</sup>, with the tablet being taken 3 h before test. Thus, subjects were assessed twice, once after single-dose, and once after week-long administration of the drug. This repeated-measures study design enabled (i) an assessment of a pharmacological effect overall, as both sessions were performed under the influence of the drug, and (ii) an assessment of putative differences between acute (single-dose) and prolonged (week-long) treatment.

**Affective state questionnaires.** To examine putative effects of the drug on subjective affective states over the course of the study, participants completed the Beck's Depression Inventory (BDI-II,<sup>69</sup>) Snaith-Hamilton Pleasure Scale (SHAPS,<sup>70</sup>) State-Trait Anxiety Inventory (STAI,<sup>71</sup>) and the Positive and Negative Affective Scale (PANAS,<sup>72</sup>) on two different occasions: (i) pre-drug, day 1; (ii) peak drug, day 7.

**Experimental task.** To examine differences in learning from success and failure, we used a modified version of a gambling card game<sup>7</sup>, in which subjects' goal was to maximize monetary wins and minimize monetary losses.

The game consisted of 180 trials, divided into three 60-trial blocks. On each trial (Fig. 1a), subjects were shown with which one of three possible decks (each designated by distinct colour and pattern) they will be playing. After a short interval (2 to 5 s, uniformly distributed), the computer drew a number between 1 and 9, and participants had up to 2.5 s to choose whether they wanted to gamble that the number, which they are about to draw, will be higher than the computer drawn number. If participants chose to gamble, they won £5 if the number that they drew was higher than the computer's number, and they lost £5 if it was lower (as well as in half of the trials in which the numbers were equal). If subjects opted to decline the gamble, they won/lost with a fixed 50% known probability. On such trials, the outcome was not shown to participants. Not making any choice always resulted in a loss. Feedback was provided 700 ms following each choice and consisted of a '+£5' (win), '-£5' (loss), or '+£5 / -£5' (win or loss, 50% probability) visual symbol. The drawn number was not shown. Subjects were told that each of the three decks contained a different proportion of high and low

numbers, and they could learn by trial and error about each of the decks' likelihood of success.

Unbeknownst to participants, one deck contained a uniform distribution of numbers between 1 and 9 ('even deck'), one deck contained more 1's than other numbers ('low deck'), making gambles 30% less likely to succeed, and one deck contained more 9's than other numbers ('high deck'), making gambles 30% more likely to succeed. In the first 15 trials, the computer drew the numbers 4, 5, and 6 three times each, and the other numbers once each. To ensure that all participants gambled in approximately 50% of trials, the numbers that the computer drew three times each were increased by one (e.g., [4, 5, 6] to [5, 6, 7]), in each subsequent set of 15 trials, if subjects took two thirds or more of the gambles against these numbers in the previous 15 trials, or decreased by one if participants took a third or less of the gambles. Participants' decks were pseudorandomly ordered while ensuring that the three decks were matched against similar computers' numbers and that no deck appeared in successive trials more than the other decks.

On both sessions, the game was identical, with the only difference being subjects played with distinct sets of three decks, indicated by different colours, where colour order and colour-associated win probability randomly varied across participants (Fig. 1b). Subjects were informed that the decks from session II were entirely unrelated to the ones from session I, and they had to learn about the novel decks anew.

To familiarize participants with the basic structure of the task, subjects, on both sessions, performed a 60-trial training block with an 'even' deck, where visual feedback indicated the number that participants drew.

**Logistic regression analysis.** We fitted a trial-by-trial logistic regression model to subjects' decisions:

$$p(c_t = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t})}}, \quad (1)$$

where a subject either accepted ( $c_t = 1$ ) or declined ( $c_t = 0$ ) a gamble on trial  $t$ . Here,  $x_{1t}$  is the computer number, scaled to range between  $-1$  (for number 9) and  $1$  (for number 1).  $x_{2t}$  represents cumulative success, reflecting, for the deck played with on trial  $t$ , the sum of previous positive outcomes, computed as  $+1$  multiplied by the computer's number against which it was received.  $x_{3t}$  represents cumulative failure, reflecting, for the deck played with on trial  $t$ , the sum of previous negative outcomes, computed as  $-1$  multiplied by  $(10 - \text{computer's number})$ , against which it was received. The multiplications by the computer's number reflect the fact that a win against a higher computer number provides stronger evidence in favour of a deck, and in a similar vein, a loss against a lower computer number provides stronger evidence against a deck. Thus, we refer to regressors as cumulative success/failure, instead of merely cumulative wins/losses. To adequately compare effect sizes between coefficients,  $x_{2t}$  and  $x_{3t}$  were range-normalized between  $-1$  and  $1$ . Positive coefficients for the first predictor ( $\beta_1$ ) indicate that subjects were more likely to gamble against lower computer numbers. Positive coefficients for the second predictor ( $\beta_2$ ) indicate that subjects were more likely to gamble given a deck with which they had experienced more cumulative success. Positive coefficients for the third predictor ( $\beta_3$ ) indicate that subjects were more likely to decline a gamble given a deck with which they had experienced more cumulative failure.  $\beta_0$  is the intercept.

Note that the regression did not converge for one subject on session I, thus we discarded this data from the group analysis.

**Computational modelling**

*Model space.* To account for the precise mechanisms that guided learning from reward and punishment, we compared a set of computational learning models in terms of how well each model explained subjects' choices. Note that although our task involves gambling, it contrasts with typical risky decision-making paradigms (e.g. 54,73) in that decision variables need to be learned by trial and error. Moreover, the typical approach in risky decision-making studies for estimating a utility function is not suitable here, since gains and losses have only one possible size. Thus, we modelled the data using a variety of reinforcement learning models, which have been shown previously to adequately capture risk sensitivity in the context of trial-and-error learning<sup>7,10</sup>. In all models, the probability of taking a gamble was modelled by applying a logistic function to a term that represented available evidence.

Model 1 ('gambling bias') and model 2 ('gambling bias & computer number') are oblivious to previous experience with the decks, and do not assume any learning to occur.

Here, model 1 computes the evidence as:

$$\beta', \quad (2)$$

where  $\beta'$  is a gambling bias parameter, determining a subject's general propensity to gamble, thus allowing the model to favour either gambling or declining to begin with.

Model 2 computes the evidence as:

$$\beta' + \beta'' N_t, \quad (3)$$

where  $N$  is the computer drawn number at trial  $t$ , scaled to range between  $-1$  (for

number 9) and  $1$  (for number 1), equivalent to the logistic regression, and  $\beta''$  is an inverse temperature parameter, determining the strength, with which the computer's number is determining a decision to gamble.

Model 3 ('Q-learning') learns the expected outcome of gambles with each deck  $d$  as follows:

$$Q_{t+1}^d = Q_t^d + \eta \delta_t, \quad (4)$$

where

$$\delta_t = r_t - Q_t^d \quad (5)$$

is an outcome prediction error, reflecting the difference between the actual ( $r_t$ ) and the expected ( $Q_t^d$ ) outcome of a gamble (initialized as  $Q_0^d = 0$ ).  $r_t = 1$  represents a win, and  $r_t = -1$  represents a loss, and  $\eta$  is a learning rate parameter that weights the influence of prediction errors on subsequent expectations. Model 3 then computes the evidence as:

$$\beta' + \beta'' N_t + \beta''' Q_t^d, \quad (6)$$

where  $\beta'''$  is a free parameter that determines the strength, with which choices are directed towards higher Q-value options.

In contrast to the previous model, model 4 ('adjusted Q-learning') computes prediction errors with respect to expectations that additionally factor in the computer's number:

$$\delta_t = r_t - Q_t^d - N_t, \quad (7)$$

which means the model learns more from more surprising outcomes, i.e., from win outcomes of gambles against higher numbers, and from loss outcomes of gambles against lower numbers.

Based on prior work<sup>7,11</sup>, we assumed subjects would learn at a different rate from successful, i.e., reward, and unsuccessful gambles, i.e., punishment. In contrast to the gambling bias parameter ( $\beta'$ ) that was included in all models, allowing them to favour either gambling or declining to begin with, an asymmetric learning bias can make such a tendency evolve with learning over time.

To this end, model 5 ('asymmetric Q-learning') and model 6 ('adjusted & asymmetric Q-learning') incorporate two distinct learning rate parameters ( $\eta^+$  &  $\eta^-$ ), that allow learning at a different rate from different outcome types, i.e., from wins:

$$Q_{t+1}^d = Q_t^d + \eta^+ \delta_t, \quad (8)$$

and from losses:

$$Q_{t+1}^d = Q_t^d + \eta^- \delta_t \quad (9)$$

Note that a model with different positive and negative learning rates for each deck could not be estimated due to the number of outcomes subjects observed varying substantially across decks and outcome type, such that not all subjects observed both positive and negative outcomes for each of the decks. Thus, in accordance with our earlier work using an equivalent task<sup>7</sup>, we assumed the same two learning rates characterized learning about all decks. We acknowledge a limitation of this approach is that learning rate estimation is more heavily influenced by trials from the high, followed by the even and then the low deck, as subjects gambled more often with better decks and consequently observed more outcomes from which they could learn.

*Model fitting.* To fit the parameters of the different models to subjects' decisions, we used an iterative hierarchical expectation-maximization procedure across the entire sample, separately for each session<sup>56,74</sup>. We sampled  $10^5$  random settings of the parameters from predefined prior distributions. Then, we computed the likelihood of observing subjects' choices given each setting and used the computed likelihoods as importance weights to re-fit the parameters of the prior distributions. These steps were repeated iteratively until model evidence ceased to increase. To derive the best-fitting parameters for each individual subject, we computed a weighted mean of the final batch of parameter settings, in which each setting was weighted by the likelihood it assigned to the individual subject's decisions. Note that the hierarchical fitting procedure, including all priors, were applied to the entire sample without distinguishing between SSRI and placebo subjects. This ensured that the parameter estimates, at the level of individual subjects, were mutually independent given the shared prior, rendering it appropriate to assess between-group differences. Learning rate parameters ( $\eta$ ,  $\eta^+$  &  $\eta^-$ ) were modelled with Beta distributions (initialized with shape parameters  $a = 1$  and  $b = 1$ ). The gambling bias parameter ( $\beta'$ ) was modelled with a normal distribution (initialized with  $\mu = 0$  and  $\sigma = 1$ ), and inverse temperature parameters ( $\beta''$  &  $\beta'''$ ) were modelled with Gamma distributions (initialized with  $\kappa = 1$ ,  $\theta = 1$ ).

*Model comparison.* We compared between models in terms of how well each model accounted for subjects' choices by means of the integrated Bayesian Information Criterion (iBIC<sup>56,75</sup>). Here, we estimated the evidence in favour of each model ( $\lambda$ ) as the mean likelihood of the model given  $10^5$  random parameter settings drawn from the fitted group-level priors. We then computed the iBIC by penalizing the model evidence to account for model complexity as follows:  $iBIC = -2 \ln \lambda + \kappa \ln n$ , where  $\kappa$  is the number of fitted parameters, and  $n$  is the total number of subject

choices used to compute the likelihood. Lower iBIC values indicate a more parsimonious model fit.

**Statistics and reproducibility.** Drug effects were assessed using repeated measures analyses of variance (rm-ANOVA) and independent samples *t*-tests. Our sample size ( $n = 66$ ) was similar to related studies related using a comparable pharmacological protocol, e.g.,<sup>26,28</sup>. We did not attempt to reproduce the pharmacological results. However, the results of our study on learning asymmetries replicate previous findings using an identical cognitive task<sup>7</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The data analysed during this study are available on GitHub ([https://github.com/jmichely/ssri\\_asymmetric\\_learning](https://github.com/jmichely/ssri_asymmetric_learning)) and Zenodo (<https://zenodo.org/badge/latestdoi/262273939>). Supplementary Data 1 contains all source data underlying the graphs and charts presented in the main figures of the manuscript.

### Code availability

The custom computer code used to generate the reported results are available GitHub ([https://github.com/jmichely/ssri\\_asymmetric\\_learning](https://github.com/jmichely/ssri_asymmetric_learning)) and Zenodo (<https://zenodo.org/badge/latestdoi/262273939>).

Received: 19 November 2020; Accepted: 8 July 2022;

Published online: 12 August 2022

### References

- Skinner, B. F. *The behavior of organisms: an experimental analysis*. (Appleton-Century, 1938).
- Kubaneck, J., Snyder, L. H. & Abrams, R. A. Reward and punishment act as distinct factors in guiding behavior. *Cognition* **139**, 154–167 (2015).
- Palminteri, S. & Pessiglione, M. In *Decision Neuroscience* (eds. Dreher, J. C. & Tremblay, L.) (Academic Press, 2017).
- Wachter, T., Lungu, O. V., Liu, T., Willingham, D. T. & Ashe, J. Differential effect of reward and punishment on procedural learning. *J. Neurosci.* **29**, 436–443 (2009).
- Monosov, I. E. & Hikosaka, O. Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex. *J. Neurosci.* **32**, 10318–10330 (2012).
- Palminteri, S. et al. Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron* **76**, 998–1009 (2012).
- Eldar, E., Hauser, T. U., Dayan, P. & Dolan, R. J. Striatal structure and function predict individual biases in learning to avoid pain. *Proc. Natl Acad. Sci. USA* **113**, 4812–4817 (2016).
- Huys, Q. J. et al. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol.* **7**, e1002028 (2011).
- Guitart-Masip, M., Duzel, E., Dolan, R. & Dayan, P. Action versus valence in decision making. *Trends Cogn. Sci.* **18**, 194–202 (2014).
- Niv, Y., Edlund, J. A., Dayan, P. & O’Doherty, J. P. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J. Neurosci.* **32**, 551–562 (2012).
- Gershman, S. J. Do learning rates adapt to the distribution of rewards. *Psychon. Bull. Rev.* **22**, 1320–1327 (2015).
- Caze, R. D. & van der Meer, M. A. Adaptive properties of differential learning rates for positive and negative outcomes. *Biol. Cyber.* **107**, 711–719 (2013).
- Pulcu, E. & Browning, M. Affective bias as a rational response to the statistics of rewards and punishments. *Elife* **6**, <https://doi.org/10.7554/eLife.27879> (2017).
- Palminteri, S., Khamassi, M., Joffily, M. & Coricelli, G. Contextual modulation of value signals in reward and punishment learning. *Nat. Commun.* **6**, 8096 (2015).
- Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S. & Palminteri, S. Behavioural and neural characterization of optimistic reinforcement learning. *Nat. Hum. Behav.* **1**, 0067 (2017).
- Murphy, F. C., Michael, A., Robbins, T. W. & Sahakian, B. J. Neuropsychological impairment in patients with major depressive disorder: the effects of feedback on task performance. *Psychol. Med.* **33**, 455–467 (2003).
- Eshel, N. & Roiser, J. P. Reward and punishment processing in depression. *Biol. Psychiatry* **68**, 118–124 (2010).
- 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).
- Bodi, N. et al. Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson’s patients. *Brain* **132**, 2385–2395 (2009).
- Cools, R. et al. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *J. Neurosci.* **29**, 1538–1543 (2009).
- Palminteri, S. et al. Pharmacological modulation of subliminal learning in Parkinson’s and Tourette’s syndromes. *Proc. Natl Acad. Sci. USA* **106**, 19179–19184 (2009).
- Chamberlain, S. R. et al. Neurochemical modulation of response inhibition and probabilistic learning in humans. *Science* **311**, 861–863 (2006).
- Cools, R., Robinson, O. J. & Sahakian, B. Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. *Neuropsychopharmacology* **33**, 2291–2299 (2008).
- Tanaka, S. C. et al. Serotonin affects association of aversive outcomes to past actions. *J. Neurosci.* **29**, 15669–15674 (2009).
- den Ouden, H. E. et al. Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* **80**, 1090–1100 (2013).
- Skandali, N. et al. Dissociable effects of acute SSRI (escitalopram) on executive, learning and emotional functions in healthy humans. *Neuropsychopharmacology* **43**, 2645–2651 (2018).
- Palminteri, S., Clair, A. H., Mallet, L. & Pessiglione, M. Similar improvement of reward and punishment learning by serotonin reuptake inhibitors in obsessive-compulsive disorder. *Biol. Psychiatry* **72**, 244–250 (2012).
- Scholl, J. et al. Beyond negative valence: 2-week administration of a serotonergic antidepressant enhances both reward and effort learning signals. *PLoS Biol.* **15**, e2000756 (2017).
- Faulkner, P. & Deakin, J. F. The role of serotonin in reward, punishment and behavioural inhibition in humans: insights from studies with acute tryptophan depletion. *Neurosci. Biobehav. Rev.* **46**, 365–378 (2014).
- van Donkelaar, E. L. et al. Mechanism of acute tryptophan depletion: is it only serotonin? *Mol. Psychiatry* **16**, 695–713 (2011).
- Crockett, M. J. et al. Converging evidence for central 5-HT effects in acute tryptophan depletion. *Mol. Psychiatry* **17**, 121–123 (2012).
- Guitart-Masip, M. et al. Differential, but not opponent, effects of L-DOPA and citalopram on action learning with reward and punishment. *Psychopharmacology* **231**, 955–966 (2014).
- Burghardt, N. S. & Bauer, E. P. Acute and chronic effects of selective serotonin reuptake inhibitor treatment on fear conditioning: implications for underlying fear circuits. *Neuroscience* **247**, 253–272 (2013).
- Burghardt, N. S., Sullivan, G. M., McEwen, B. S., Gorman, J. M. & LeDoux, J. E. The selective serotonin reuptake inhibitor citalopram increases fear after acute treatment but reduces fear with chronic treatment: a comparison with tianeptine. *Biol. Psychiatry* **55**, 1171–1178 (2004).
- Bari, A. et al. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology* **35**, 1290–1301 (2010).
- Baldinger, P. et al. Regional differences in SERT occupancy after acute and prolonged SSRI intake investigated by brain PET. *Neuroimage* **88**, 252–262 (2014).
- Correia, P. A. et al. Transient inhibition and long-term facilitation of locomotion by phasic optogenetic activation of serotonin neurons. *Elife* **6**, <https://doi.org/10.7554/eLife.20975> (2017).
- Gutierrez, M. & Abramowitz, W. Steady-state pharmacokinetics of citalopram in young and elderly subjects. *Pharmacotherapy* **20**, 1441–1447 (2000).
- Godlewska, B. R. & Harmer, C. J. Cognitive neuropsychological theory of antidepressant action: a modern-day approach to depression and its treatment. *Psychopharmacology*, <https://doi.org/10.1007/s00213-019-05448-0> (2020).
- Dayan, P. Twenty-five lessons from computational neuromodulation. *Neuron* **76**, 240–256 (2012).
- Olivier, B. Serotonin: a never-ending story. *Eur. J. Pharm.* **753**, 2–18 (2015).
- Crockett, M. J. & Cools, R. Serotonin and aversive processing in affective and social decision-making. *Curr. Opin. Behav. Sci.* **5**, 64–70 (2015).
- Moustafa, A. A., Gluck, M. A., Herzallah, M. M. & Myers, C. E. The influence of trial order on learning from reward vs. punishment in a probabilistic categorization task: experimental and computational analyses. *Front Behav. Neurosci.* **9**, 153 (2015).
- Macoveanu, J. Serotonergic modulation of reward and punishment: evidence from pharmacological fMRI studies. *Brain Res* **1556**, 19–27 (2014).
- Maya Vetencourt, J. F. et al. The antidepressant fluoxetine restores plasticity in the adult visual cortex. *Science* **320**, 385–388 (2008).
- Karpova, N. N. et al. Fear erasure in mice requires synergy between antidepressant drugs and extinction training. *Science* **334**, 1731–1734 (2011).



47. Krishnan, V. & Nestler, E. J. Linking molecules to mood: new insight into the biology of depression. *Am. J. Psychiatry* **167**, 1305–1320 (2010).
48. Hauser, T. U., Eldar, E. & Dolan, R. J. Neural Mechanisms of Harm-Avoidance Learning: A Model for Obsessive-Compulsive Disorder? *JAMA Psychiatry* **73**, 1196–1197 (2016).
49. Campbell-Meiklejohn, D. et al. Serotonin and dopamine play complementary roles in gambling to recover losses. *Neuropsychopharmacology* **36**, 402–410 (2011).
50. Macoveanu, J. et al. Effects of selective serotonin reuptake inhibition on neural activity related to risky decisions and monetary rewards in healthy males. *Neuroimage* **99**, 434–442 (2014).
51. Macoveanu, J. et al. Playing it safe but losing anyway—serotonergic signaling of negative outcomes in dorsomedial prefrontal cortex in the context of risk-aversion. *Eur. Neuropsychopharmacol.* **23**, 919–930 (2013).
52. Hieronymus, F., Lisinski, A., Nilsson, S. & Eriksson, E. Efficacy of selective serotonin reuptake inhibitors in the absence of side effects: a mega-analysis of citalopram and paroxetine in adult depression. *Mol. Psychiatry* **23**, 1731–1736 (2018).
53. Harmer, C. J., Duman, R. S. & Cowen, P. J. How do antidepressants work? New perspectives for refining future treatment approaches. *Lancet Psychiatry* **4**, 409–418 (2017).
54. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. A computational and neural model of momentary subjective well-being. *Proc. Natl Acad. Sci. USA* **111**, 12252–12257 (2014).
55. Rutledge, R. B., Skandali, N., Dayan, P. & Dolan, R. J. Dopaminergic Modulation of Decision Making and Subjective Well-Being. *J. Neurosci.* **35**, 9811–9822 (2015).
56. Eldar, E., Roth, C., Dayan, P. & Dolan, R. J. Decodability of Reward Learning Signals Predicts Mood Fluctuations. *Curr. Biol.* **28**, 1433–1439.e1437 (2018).
57. Eldar, E. & Niv, Y. Interaction between emotional state and learning underlies mood instability. *Nat. Commun.* **6**, 6149 (2015).
58. Otto, A. R. & Eichstaedt, J. C. Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. *PLoS One* **13**, e0206923 (2018).
59. Villano, W. J., Otto, A. R., Ezie, C. E. C., Gillis, R. & Heller, A. S. Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. *J. Exp. Psychol. Gen.* **149**, 1755–1766 (2020).
60. Eldar, E., Rutledge, R. B., Dolan, R. J. & Niv, Y. Mood as Representation of Momentum. *Trends Cogn. Sci.* **20**, 15–24 (2016).
61. Sharot, T., Guitart-Masip, M., Korn, C. W., Chowdhury, R. & Dolan, R. J. How dopamine enhances an optimism bias in humans. *Curr. Biol.* **22**, 1477–1481 (2012).
62. Sharot, T., Riccardi, A. M., Raio, C. M. & Phelps, E. A. Neural mechanisms mediating optimism bias. *Nature* **450**, 102–105 (2007).
63. Garrett, N. et al. Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. *Front Hum. Neurosci.* **8**, 639 (2014).
64. Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R. & Dolan, R. J. Depression is related to an absence of optimistically biased belief updating about future life events. *Psychol. Med.* **44**, 579–592 (2014).
65. Taylor, M. J., Freemantle, N., Geddes, J. R. & Bhagwagar, Z. Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Arch. Gen. Psychiatry* **63**, 1217–1223 (2006).
66. Michely, J., Eldar, E., Martin, I. M. & Dolan, R. J. A mechanistic account of serotonin's impact on mood. *Nat. Commun.* **11**, 2335 (2020).
67. Michely, J., Martin, I. M., Dolan, R. J. & Hauser, T. U. Boosting serotonin increases information gathering by reducing subjective cognitive costs. *bioRxiv* <https://doi.org/10.1101/2021.12.08.471843> (2021).
68. Noble, S. & Benfield, P. Citalopram: A Review of its Pharmacology, Clinical Efficacy and Tolerability in the Treatment of Depression. *CNS Drugs* **8**, 410–431 (1997).
69. Beck, A. T., Steer, R. A. & Brown, G. K. *Manual for the Beck Depression Inventory-II*. (Psychological Corporation, 1996).
70. Snaith, R. P. et al. A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *Br. J. Psychiatry* **167**, 99–103 (1995).
71. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. *Manual for the State-Trait Anxiety Inventory*. (Consulting Psychologists Press, 1983).
72. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988).
73. Sokol-Hessner, P. et al. Thinking like a trader selectively reduces individuals' loss aversion. *Proc. Natl Acad. Sci. USA* **106**, 5035–5040 (2009).
74. Bishop, C. M. *Pattern Recognition and Machine Learning*. (Springer, 2006).
75. Huys, Q. J. et al. Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol.* **8**, e1002410 (2012).

## Acknowledgements

J.M. was supported by a fellowship from the German Research Foundation (MI 2158/1-1) and is participant in the BIH Charité (Junior) (Digital) Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin, and the Berlin Institute of Health at Charité (BIH). R.J.D. holds a Wellcome Trust Investigator award (098362/Z/12/Z). The Max Planck UCL Centre for Computational Psychiatry and Ageing Research is a joint initiative supported by the Max Planck Society and University College London. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust (091593/Z/10/Z).

## Author contributions

J.M. and E.E. designed the experiments. J.M. and I.M.M. performed the experiments. J.M., E.E., A.E. and R.J.D. analysed and interpreted the data. J.M., E.E. and R.J.D. wrote the paper.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03690-5>.

**Correspondence** and requests for materials should be addressed to Jochen Michely.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Enzo Tagliacucchi and George Inglis. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022