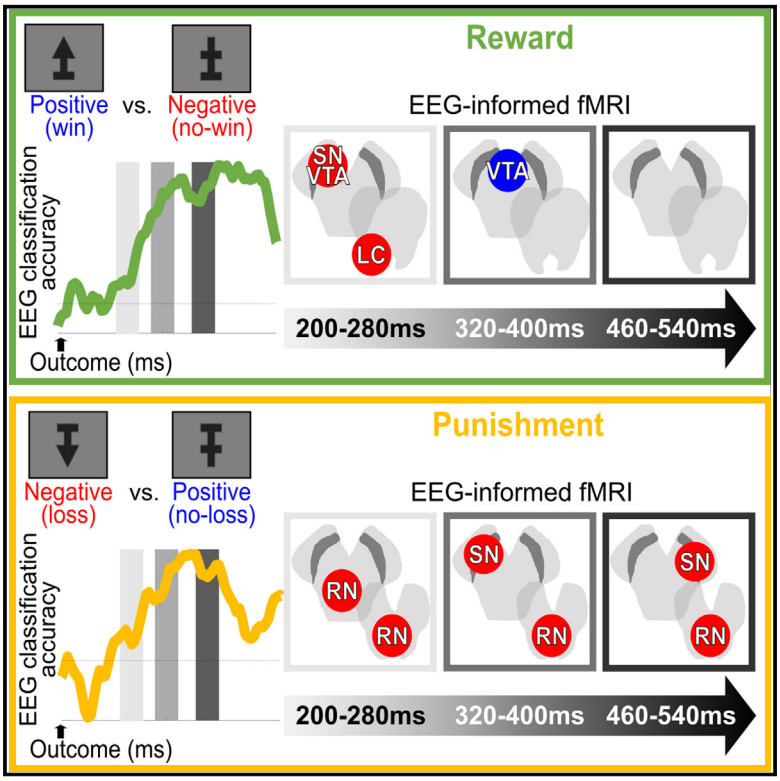
Cell Reports

Article

Distinct spatiotemporal brainstem pathways of outcome valence during reward- and punishment-based learning

Graphical abstract



Highlights

- Spatiotemporal characterization of outcome valence signals using EEG-fMRI fusion
- Distinct brainstem pathways of outcome valence in reward and punishment learning
- Broad overlap of cortical outcome valence signals in reward and punishment learning
- Behavioral choices shaped by the coupling of brainstem nuclei and other areas

Authors

Joana Carvalheiro, Marios G. Philiastides

Correspondence

joana.carvalheiro@glasgow.ac.uk (J.C.), marios.philiastides@glasgow.ac.uk (M.G.P.)

In brief

Carvalheiro and Philiastides find, using EEG-fMRI, distinct spatiotemporal brainstem pathways of outcome valence responses during reward- and punishment-based learning. Additionally, the coupling of these distinct brainstem pathways with other brain areas modulates behavior. These findings provide a comprehensive spatiotemporal account of how the brain signals positive and negative outcomes in reward vs. punishment contexts.

Carvalheiro & Philiastides, 2023, Cell Reports *42*, 113589
December 26, 2023
https://doi.org/10.1016/j.celrep.2023.113589

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Distinct spatiotemporal brainstem pathways of outcome valence during reward- and punishment-based learning

Joana Carvalheiro[1,2,3,*] and Marios G. Philiastides[1,2,*]
[1]School of Psychology and Neuroscience, University of Glasgow, Glasgow G12 8QB, UK
[2]Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow G12 8QB, UK
[3]Lead contact
*Correspondence: joana.carvalheiro@glasgow.ac.uk (J.C.), marios.philiastides@glasgow.ac.uk (M.G.P.)
https://doi.org/10.1016/j.celrep.2023.113589

SUMMARY

Learning to seek rewards and avoid punishments, based on positive and negative choice outcomes, is essential for human survival. Yet, the neural underpinnings of outcome valence in the human brainstem and the extent to which they differ in reward and punishment learning contexts remain largely elusive. Here, using simultaneously acquired electroencephalography and functional magnetic resonance imaging data, we show that during reward learning the substantia nigra (SN)/ventral tegmental area (VTA) and locus coeruleus are initially activated following negative outcomes, while the VTA subsequently re-engages exhibiting greater responses for positive than negative outcomes, consistent with an early arousal/avoidance response and a later value-updating process, respectively. During punishment learning, we show that distinct raphe nucleus and SN subregions are activated only by negative outcomes with a sustained post-outcome activity across time, supporting the involvement of these brainstem subregions in avoidance behavior. Finally, we demonstrate that the coupling of these brainstem structures with other subcortical and cortical areas helps to shape participants' serial choice behavior in each context.

INTRODUCTION

Humans, and other animals, constantly use positive and negative feedback to adjust their behavior toward maximizing rewards and minimizing punishments. Positive outcomes (reward or omission of punishment) increase the likelihood of repeating the same choice, and negative outcomes (omission of reward or punishment) increase the likelihood of avoiding that choice in the future. While distinct brainstem subregions have been implicated in reinforcement learning,[1–4] especially in invasive non-human animal studies, a full spatiotemporal account of the human brainstem pathways associated with outcome valence during reward and punishment learning is still lacking. Human functional magnetic resonance imaging (fMRI) studies often lack the necessary sensitivity to isolate small subcortical structures but, more critically, lack the temporal resolution to identify the relative timing with which subcortical outcome valence signals might emerge. Moreover, a direct comparison of these temporal dynamics during human reward vs. punishment learning—under the same experimental setting—is also lacking.

During reward learning, animal electrophysiological studies identified fast dopaminergic responses in substantia nigra/ventral tegmental area (SN/VTA) complex, with increased activity in response to unexpected positive outcomes and decreased activity in response to unexpected negative outcomes.[2,5–9] Non-human animal studies also implicated noradrenergic neurons in the locus coeruleus (LC) in reward learning[1,10–12] by enhancing arousal in response to salient events, such as omissions of expected rewards, and redirecting attention toward negative outcomes and facilitate behavioral change.[1,11,13–15] In humans, the evidence for outcome valence signals in SN/VTA and LC is scarcer but largely consistent with non-human animal studies.[16–19]

The role of brainstem subregions in signaling outcome valence in punishment learning is more debatable. For example, while some non-human animal studies argued that outcome valence is encoded by SN/VTA neurons in a similar manner as in reward learning (i.e., increases/decreases following positive/negative outcomes),[7] others point to mainly positive activations of SN/VTA neurons following negative punishing outcomes,[6,20] possibly via distinct neuronal subpopulations.[2,5,7,21–24] Similarly, raphe nucleus (RN) serotoninergic neurons have been implicated in punishment learning, by responding to negative outcomes and mediating avoidance behavior,[3,25–28] likely via their projections to other subcortical and cortical areas.[29–31] Indirect evidence for the role of RN in punishment learning also comes from human pharmacological studies using serotoninergic agonists/antagonists.[32–38]

While the SN/VTA, LC, and RN have been implicated in reward and punishment learning, the extent to which they encode different outcome valence signals that cascade rapidly in time—and are thus intermixed at the level of macroscopic fMRI

activity and likely single-neuron responses as well[39]—remains unclear. Importantly, these brainstem structures broadcast widely to downstream cortical areas such that these different outcome valence signals could be independently communicated to separable cortical areas to regulate adaptive behavioral responses.[1,30,40–44] For example, using fusion of electrocenphalography (EEG) and fMRI data, Fouragnan et al.[45] showed that in reward learning, an early outcome valence system initiated an automatic alertness response following negative outcomes while in parallel downregulated activity of a later reward-related system to promote avoidance learning. Conversely, positive outcomes primarily activated the later system, consistent with a role in approach learning and value updating.

This organization of neuronal information flow offers an opportunity to first intercept and decouple these separable signals of outcome valence at the level of cortical responses (i.e., downstream of their brainstem counterparts) using high-temporal-resolution EEG measurements. Subsequently, these electrophysiological signatures can be mapped back onto distinct subcortical structures using fusion of EEG and concurrently acquired high-resolution brainstem fMRI. Importantly, this approach can be deployed on both reward- as well as punishment-based learning, within the same experimental session, to enable direct comparisons of the spatiotemporal brainstem pathways involved in each context, a critical endeavor toward a comprehensive understanding of human reinforcement learning.

Here, using simultaneously acquired EEG-fMRI, we show that there are significant differences in how outcome valence signals are encoded in the human brainstem. Critically, our findings are uniquely enabled via our EEG-fMRI fusion and not seen with a traditional stand-alone fMRI analysis.

Specifically, during reward learning, we show that the SN/VTA and LC are initially activated by negative outcomes, while a distinct VTA cluster subsequently exhibits greater responses for positive than negative outcomes, consistent with an early arousal/avoidance response and a later value-updating process, respectively. During punishment learning, RN and SN are only activated by negative outcomes and show more sustained post-outcome activity across time, supporting the role of these brainstem subregions in avoidance behavior. We corroborate these findings further by demonstrating that the coupling of these brainstem structures with other subcortical and cortical areas help shape participants' serial choice behavior.

RESULTS

Probabilistic reversal-learning task performance

We analyzed simultaneous EEG-fMRI data from 28 participants while they performed a probabilistic reversal-learning task divided into reward and punishment blocks. In the reward context, participants could win one (positive outcome) or zero (negative outcome) points, and in the punishment context, they could lose zero (positive outcome) or one (negative outcome) points. On each trial, subjects were asked to choose between two abstract symbols, which were associated with different probabilities (70% and 30%) of positive or negative outcomes, and through feedback, they had to learn to select the symbol with the highest probability of positive outcomes in the reward context and to avoid the symbol with the highest probability of negative outcomes in the punishment context (Figure 1A). However, in each block "reversals" in contingencies were introduced whereby the high/low probability was re-assigned to the opposite symbol, and subjects entered a new learning phase. Each block comprised three reversals (every 20 ± 2 trials), resulting in four learning phases.

Participants' choices tracked these reversals (Figure 1B) and were probabilistic based on expected values assigned to each symbol on individual trials (Figure S1A), in line with the principles of reinforcement learning (see STAR Methods). Choice accuracy, as well response times, did not differ between reward and punishment contexts (both $p > 0.44$; Figures 1C and 1D). As expected, participants switched their choices to a larger extent following negative than positive outcomes ($F(1,27) = 114.86$, $p < 0.001$, Figure 1E) and were also slower when making a choice following negative outcomes ($F(1,27) = 4.88$, $p = 0.036$, Figure 1F), regardless of the context (both interactions outcome valence × context, $p > 0.68$). Moreover, participants updated the expected value of the chosen stimulus (estimated using a reinforcement-learning model) to a larger extent following negative outcomes than following positive outcomes ($F(1,27) = 341.32$, $p < 0.001$, Figure 1G) in both contexts (interaction outcome valence × context, $p > 0.75$). Taken together, these findings suggest that while choice behavior is comparable across reward and punishment contexts, positive and negative outcomes contribute differently to behavioral adaptation, with the negative outcomes being the main driver of the behavioral changes observed in our task.

Temporal cascade of outcome valence signals

To identify temporally distinct neuronal signals associated with outcome valence (positive and negative) and to compare these signals across reward and punishment contexts, we used single-trial multivariate discriminant analysis of the EEG signals locked to the delivery of the outcome. This analysis was performed separately for each participant. Specifically, we estimated spatial weightings of the EEG sensors discriminating between positive vs. negative outcome trials across different outcome-locked time windows (from −100 to 850 ms relative to outcome onset)[45,46] in each of the reward and punishment contexts separately.

Applying these temporally specific spatial weights to single-trial EEG data produces a measurement of the trial-wise amplitudes of the signals discriminating outcome valence. These amplitudes (y-values; see STAR Methods, Equation 5) can be thought of as a proxy of the neuronal response variability following positive and negative outcomes, with activity common to both types of outcomes removed. Our discriminator was designed to map positive and negative outcomes to a continuum of positive and negative discriminator amplitudes, respectively. We view these amplitudes as representing a graded response to the different outcome types: large positive amplitudes are reflective of a strong response to a positive outcome, large negative amplitudes are reflective of a strong response to a negative outcome, and intermediate magnitude amplitudes are reflective of a weaker response to either outcome. While the outcome was

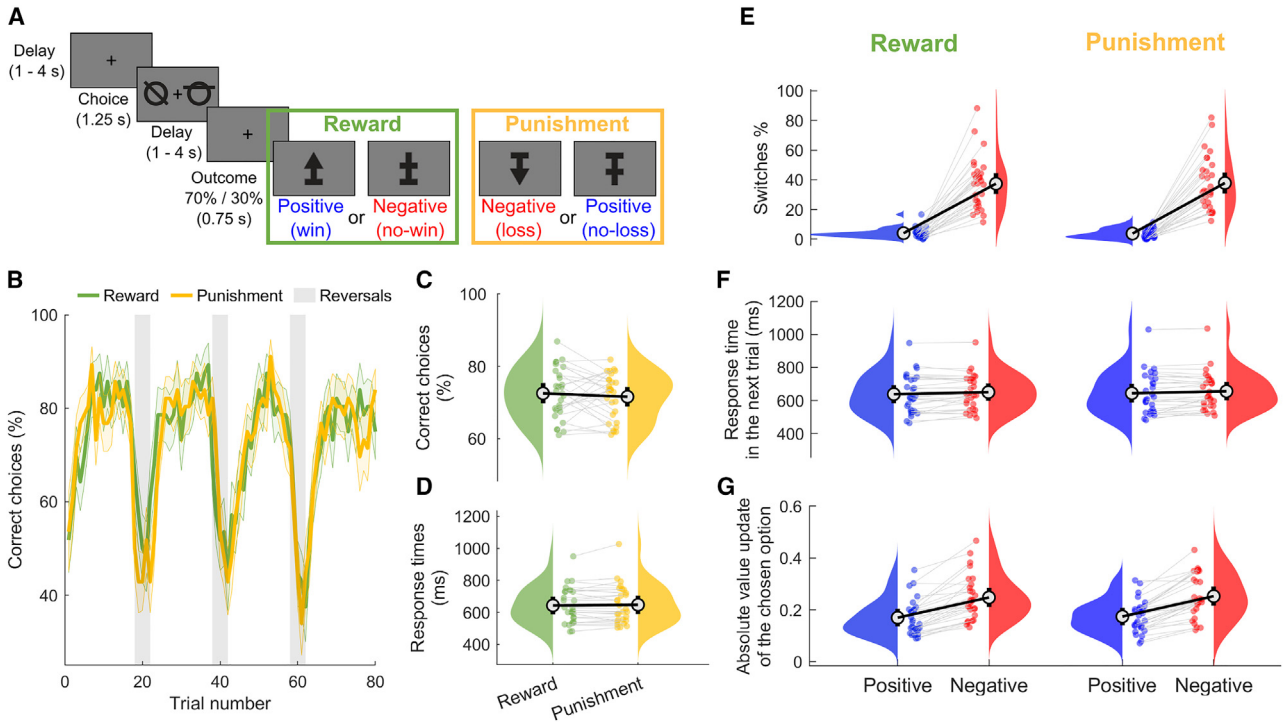


Figure 1. Experimental task and behavioral measures

(A) Schematic representation of the probabilistic reversal-learning task. On each trial, subjects had to choose between two abstract symbols carrying different probabilities (70% and 30%) of yielding positive and negative outcomes, in separate reward and punishment contexts. Once a choice was made, the outcome was revealed by using different arrows in each context. In the reward context, participants could win one (positive outcome) or zero (negative outcome) points, and in the punishment context they could lose zero (positive outcome) or one (negative outcome) points.

(B) Trial-by-trial percentage of choosing the "correct" symbol (i.e., the symbol associated with the highest probability of yielding positive outcomes) across participants in the reward and punishment contexts (n = 28 subjects). Shaded gray areas represent the "reversals" in the contingencies. Each central line represents the mean, and the filled colored areas represent the ±standard error of the mean.

(C) Mean percentage of "correct" choices in reward and punishment contexts.

(D) Mean responses times in reward and punishment contexts.

(E) Mean percentage of switches following positive and negative outcomes in reward and punishment contexts.

(F) Mean response times following positive and negative outcomes in reward and punishment contexts.

(G) Mean update of the expected value of the chosen stimulus following positive and negative outcomes in reward and punishment contexts. In (C)–(G), connected dots represent data points from the same subject; the error bars displayed on the side of the scatterplots indicate the sample mean ± standard error of the mean.

categorical (positive or negative), we view these amplitudes as representing endogenous variability in the encoding of individual trial outcomes, which we will leverage to enable the fusion of EEG with fMRI.

To quantify the discriminator's performance over time, we used the area under a receiver operating characteristic curve (that is, A_z value) with a leave-one-out trial cross-validation approach. To visualize the spatial distribution of the relevant discriminating activity, we computed forward models of this activity over time (see STAR Methods, Equation 6). Across participants, we identified a time period of significant (above chance) discriminator performance, representing a cascade of three spatiotemporally distinct EEG components discriminating between positive and negative outcomes, in both reward and punishment contexts. Specifically, a first midline component emerged between 200 and 280 ms, a second centroparietal component between 320 and 400 ms, and a third temporoparietal and far frontal component between 460 and 540 ms (Figure 2). The spatial topographies for the three components were comparable across reward and punishment contexts, suggestive of potentially similar cortical neural generators driving the relevant outcome valence signals over time. In a separate control experiment, we showed that none of these components arose due to differences in the visual properties of the outcome stimuli (see Figure S1B).

Spatiotemporal dissociation of outcome valence signals

In stand-alone fMRI, outcome valence signals would typically be identified via a categorical contrast of positive vs. negative outcome trials. Critically, however, there are two major shortcomings with this approach. Firstly, the sluggish nature of the blood-oxygen-level-dependent (BOLD) signal precludes a temporal dissociation of the relevant neural systems (i.e., activations maps are "static"), which as we demonstrated above are

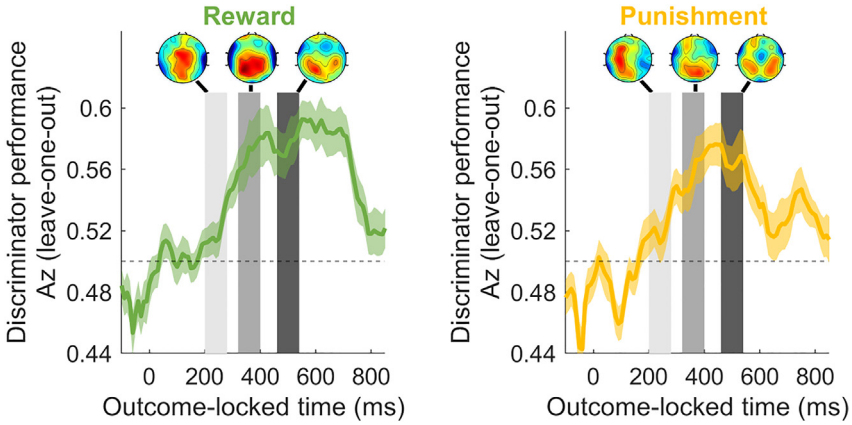


Figure 2. EEG-temporal components of outcome valence

Multivariate discriminator performance (A_z) during positive vs. negative outcome discrimination of outcome-locked EEG responses in reward (left) and punishment (right) contexts. Shaded gray areas represent the three outcome value components with spatially distinct scalp topographies (averaged across each time window). Each central line represents the mean across subjects (n = 28), and the filled colored areas represent the ±standard error of the mean. The dotted line indicates an A_z value of 0.5 (chance level).

cascading in rapid temporal succession. Secondly, conventional fMRI contrasts reveal only relative differences between categorical variables (here, positive vs. negative outcomes) and do not consider endogenous neural fluctuations within nominally identical outcomes (i.e., within positive and negative outcomes, separately) that could afford additional explanatory power in revealing the underlying neural networks (especially in smaller subcortical structures suffering from lower signal-to-noise ratio in the fMRI).

Here, we used the endogenous single-trial variability in the electrophysiological amplitudes of the three temporally distinct EEG components of outcome valence (*y*-values) to build separate parametric EEG-informed fMRI regressors for positive and negative outcomes. In other words, we leveraged the endogenous trial-by-trial variability within each of the two outcome types to obtain a better understanding of how they independently explain changes in the BOLD signal in each of the three time windows. Ultimately, this EEG-fMRI fusion controls for unspecific valence effects embedded in a conventional contrast and ensures that the relevant brain activations are now also defined in terms of their relative timing. Using this approach and brainstem-tailored fMRI sequences (Figure S2A), we aimed to test how BOLD activity in "learning-associated" subcortical regions (e.g., SN/VTA, RN, and LC) map onto each of the three outcome valence EEG components and to compare those activations between reward and punishment contexts.

Importantly, we note that the trial-by-trial variability in our EEG component amplitudes is likely influenced mainly by cortical regions near the recording electrodes and to a lesser extent by distant (e.g., subcortical) structures. The originality of our approach, however, hinges on our ability to exploit this trial-wise variability to also reveal activations from deeper subcortical structures provided their BOLD signal covaries systematically with that of the cortical sources of our EEG (e.g., by broadcasting relevant activity to dedicated cortical target sites).[1,30,41] To further validate this proposition, we collected additional data from a separate passive visual stimulation experiment and confirmed that superficial EEG activity from the visual pathway can be used to expose covarying activity in the superior colliculus in the brainstem and with additional explanatory power compared with a standard fMRI analysis (see Figure S2D). For our main experiment, we used anatomical masks to define broad regions of interest for the SN/VTA, LC, and RN and report any resulting spatial dissociations within each mask based on the location of the observed activations. To additionally characterize the spatiotemporal dynamics of the EEG cortical sources, we extended our analyses to the rest of the brain covered by our fMRI field of view. For comparison, and to further demonstrate the advantage of the EEG-fMRI fusion, we first performed standard generalized linear model (GLM) analyses in which we set categorical BOLD predictors for outcome valence, as commonly done in stand-alone fMRI studies.

Standard fMRI analysis of outcome valence

In a standard fMRI analysis, we contrasted categorical outcome regressors for positive and negative outcomes (GLM 1; STAR Methods). This analysis did not reveal any brainstem clusters with greater BOLD response for positive than negative outcomes in either the reward nor the punishment context. The inverse contrast revealed greater BOLD response to negative compared to positive outcomes in the ventral portion of right SN (without overlapping with VTA) in the reward context and in a larger cluster extending across the right SN and VTA and into the ventral portion of left SN in the punishment context (Figure 3A, see also Figure S3A and Table S1). Overlapping these clusters in reward and punishment contexts showed a similar pattern of activations only in the right ventral portion of SN (Figure 3A, see also Figure S3A). In the punishment context, we also found greater activation to negative compared to positive outcomes in median RN. These results broadly suggest that negative outcomes engage the brainstem to a larger extent than positive outcomes, in both reward and punishment contexts, consistent with the behavioral results reported above.

Assessing the categorical contrasts of positive > negative and negative > positive outcomes beyond the brainstem (across the remaining fMRI field of view, Figure S2A) revealed distributed and broadly similar brain networks across reward and punishment contexts (Figure 4A, see also Figure S3B and Table S2). Specifically, regions in which the BOLD signal was greater for positive than negative outcomes included areas of the human reward/value network such as the ventromedial prefrontal cortex, striatum, and amygdala, whereas regions in which the

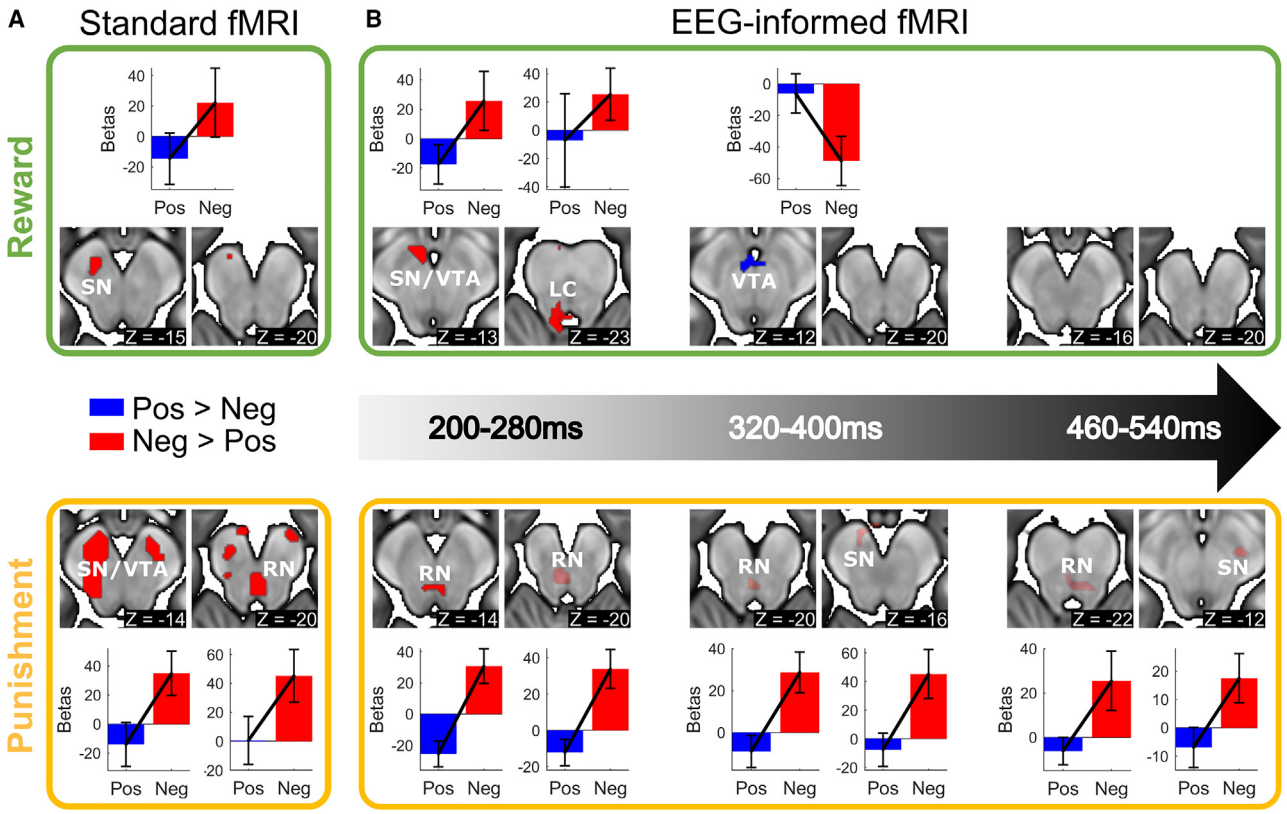


Figure 3. Spatiotemporal dissociation of outcome valence signals in the brainstem

(A) Activations in the brainstem for negative > positive outcomes in reward (upper panel) and punishment (bottom panel) contexts using standard categorical outcome regressors (all *Z* scores >2.3, cluster corrected; GLM 1, see STAR Methods; see also Figure S3A and Table S1).

(B) Activations in the brainstem for positive > negative and negative > positive outcomes in reward and punishment contexts using separate parametric regressors for positive and negative outcomes based on the EEG single-trial variability for three distinct outcome-locked time windows (all *Z* scores >2.3, cluster corrected). Opaque clusters resulted from GLMs that included EEG-informed regressors for all three time windows (GLM 2, see STAR Methods), and transparent clusters resulted from GLMs that included EEG-informed regressors for each time window separately (GLM 3, see STAR Methods; see also Figure S4A and Table S3). In (A) and (B), bars depict mean parameter estimates (i.e., betas) for the BOLD response for positive and negative outcomes within each cluster across subjects (n = 28). Error bars indicate the mean ± standard error of the mean. In each depicted slice, *Z* is the z coordinate in the standard MNI brain. SN/VTA, substantia nigra/ventral tegmental area; LC, locus coeruleus; RN, raphe nucleus.

BOLD signal was greater for negative than positive outcomes included the thalamus, insula, and prefrontal cortex. Overall, these results agree with a large body of literature reporting activations relating to contrasts of positive vs. negative outcomes.[47–50]

EEG-informed fMRI analysis of outcome valence

While the standard fMRI analysis revealed a set of activations in the brainstem for the negative > positive outcome contrast, their relative timing remains unclear. It is also possible that additional brainstem clusters might activate transiently, including some that show the opposite effect (i.e., greater response to positive than negative outcomes). In turn, these signals could be multiplexed and averaged out at the level of macroscopic BOLD activity when using a standard fMRI analysis. To obtain a more comprehensive understanding of the spatiotemporal dynamics of outcome valence in the brainstem activations, we leveraged the relevant electrophysiological components we obtained above (Figure 2) to perform an EEG-informed fMRI analysis (GLM 2; see STAR Methods). In this analysis, we used the single-trial variability in the outcome valence *y*-values derived for each of the three temporally distinct EEG components to build separate fMRI regressors for positive and negative outcomes (i.e., 6 regressors; 2 outcome types × 3 time windows). We repeated this EEG-informed fMRI analysis for each of the reward and punishment contexts separately. Finally, we tested the contrasts for positive > negative outcomes and negative > positive outcomes on these EEG-informed fMRI regressors. We note that because negative outcomes were mapped to negative *y*-values in the EEG discrimination analysis (see above), we flipped the sign of the *y*-values in the negative outcome regressors so that the contrasts between positive and negative outcome regressors remained meaningful. The GLMs described above (GLM 2) included concurrent EEG-informed regressors for the three temporally distinct EEG components and were designed specifically to detect brain activations that were stronger/unique for a specific component (that is, cannot be explained by any shared variance across regressors). However, it is possible

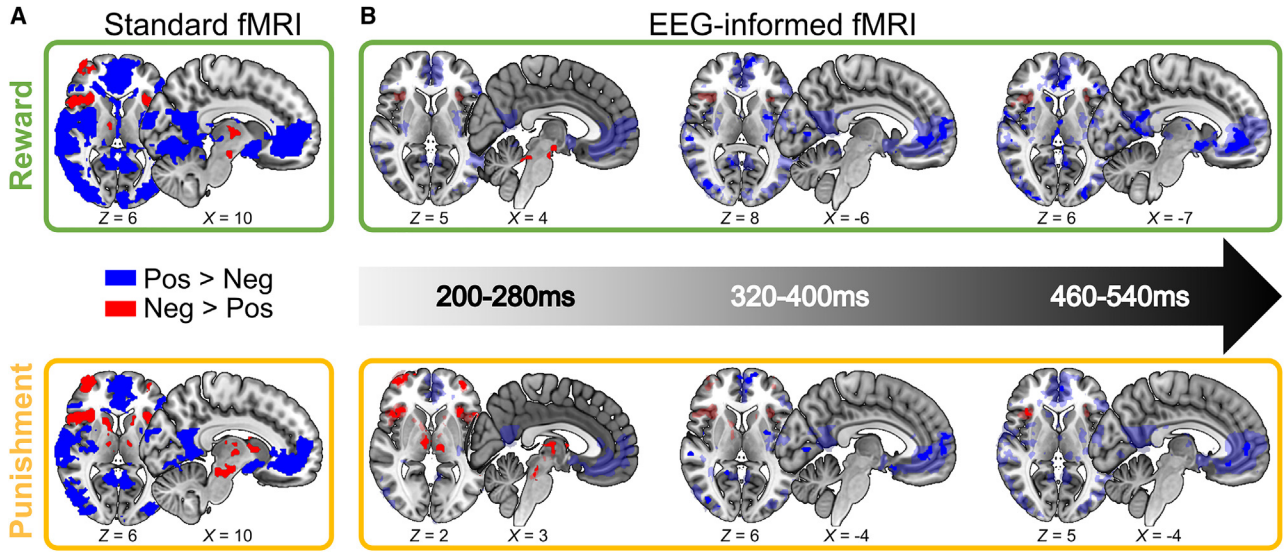


Figure 4. Spatiotemporal dissociation of outcome valence signals in the entire fMRI field of view

(A) Activations for positive > negative and negative > positive outcomes in reward (upper panel) and punishment (bottom panel) contexts using standard categorical outcome regressors (all *Z* scores >2.3, cluster corrected; GLM 1, see STAR Methods; see also Figure S3B and Table S2, n = 28 subjects).

(B) Activations for positive > negative and negative > positive outcomes in reward and punishment contexts using separate parametric regressors for positive and negative outcomes based on the EEG single-trial variability for the three time windows (all *Z* scores >2.3, cluster corrected). Opaque clusters resulted from GLMs that included the EEG-informed regressors for all three time windows (GLM 2, see STAR Methods), and transparent clusters resulted from GLMs that included EEG-informed regressors for each time window separately (GLM 3, see STAR Methods; see also Figure S4B and Table S4, n = 28 subjects). Below each depicted slice, *Z* is the z coordinate in the standard MNI brain.

that a brain region that was comparably activated across the three time windows could remain undetected in the above analysis after the shared variance in the three time windows was factored out. To address this, we also explored additional EEG-informed GLMs separately for each of the temporally distinct components (GLM 3; STAR Methods).

In the reward context, and unlike the standard fMRI analysis, we dissociated temporally distinct brainstem activation profiles for both the negative > positive and positive > negative contrasts (GLM 2). Specifically, we identified a cluster in LC and a cluster encompassing VTA and the ventral portion of right SN exhibiting greater response for negative compared with positive outcomes in the first time window, with a distinct cluster in VTA exhibiting a greater response for positive compared with negative outcomes in the second time window (Figure 3B and Table S3). We did not find any significant activations in either contrast in the third time window. Additional separate analyses for each of the temporally distinct components (GLM 3) did not show any significant activations in the brainstem regions of interest in any of the three time windows, suggesting that the activations described above for the reward context were transient and time window specific. Importantly, the early activations in LC and later activations to positive > negative outcomes in VTA were not detected in the standard fMRI analysis. These results further confirm that the endogenous variability in our EEG-derived measure of outcome valence can offer additional explanatory power over and above standard categorical contrasts in fMRI analysis. Taken together, these findings are consistent with an early automatic alertness response to negative outcomes driven by LC and SN, followed by a later VTA response likely involved in updating value and driving reward-based learning.

In the punishment context, a cluster encompassing part of dorsal and median RN showed greater response for negative compared with positive outcomes in the first time window (Figure 3B, opaque clusters); no clusters survived correction in the other time windows (GLM 2). Additional separate analysis for each of the temporally distinct components (GLM 3) revealed a median RN cluster that showed greater response for negative than positive outcomes across the three time windows (Figure 3B, transparent clusters). We also found SN clusters (not overlapping with the VTA) showing greater response for negative than positive outcomes across the second and third time windows, albeit with different spatial locations (Figure 3B, transparent clusters). Specifically, we found a cluster in the right ventral portion of SN in the second time window and a separate cluster in the left dorsal portion of SN in the third time window (Table S3). These results suggest a quick engagement of the RN to negative outcomes, which persists throughout the outcome period, along with a later involvement of distinct portions of the SN (relative to reward learning) consistent with a role of these subregions in avoidance learning.

Whereas the reward context was characterized by transient LC and SN/VTA responses to negative and negative/positive outcomes (respectively), the punishment context was characterized by more sustained responses in the RN and distinct SN subdivisions only to negative outcomes. To further test the extent to which the brainstem responded similarly/differently to outcome valence in reward vs. punishment contexts,

we extracted the parameter estimates (i.e., betas) from each cluster and compared them between contexts. We found that the pattern of activations for positive vs. negative outcomes within each brainstem cluster was different between contexts, suggesting that the clusters reported above were primarily driven by the context from which they were derived (Figure S4A).

While the main focus of this work was to offer a spatiotemporal account of the neural correlates of outcome valence in the human brainstem, we additionally inspected the EEG-informed fMRI positive vs. negative contrasts in the remaining field of view afforded to us by our fMRI sequence (Figure S2A). We used these results to obtain a broad view on how brainstem signatures of outcome valence are likely broadcasted onto the cortex. Although we observed stark differences in the representation of outcome valence in the brainstem across reward and punishment contexts, other brain areas exhibited broadly similar spatiotemporal activation patterns across the two contexts (Figure 4B, opaque clusters; see also Figures S4B and S4C and Table S4). Specifically, in the first time window, we identified brain regions with greater responses for negative than positive outcomes (GLM 2) and that have been implicated in salience monitoring and negative outcome processing,[48,51] such as insula and prefrontal areas (Figure 4B, opaque clusters). During the second and third time windows, we identified a network of regions that showed greater response for positive than negative outcomes (GLM 2) and that comprised mainly regions of the human valuation system,[52] such as the striatum and ventromedial prefrontal cortex (Figure 4B, opaque clusters). There were additional activations in the punishment context, mainly in the first time window.

Additionally, separate GLMs for each the temporally distinct components (GLM 3) further confirmed broadly similar brain networks for outcome valence in reward and punishment contexts across the three time windows, including insula for negative > positive outcomes and ventromedial prefrontal cortex and striatum for positive > negative outcomes (Figure 4B, transparent clusters, see also Table S4). Overall, our results suggest that while brainstem pathways implicated in reward- and punishment-based learning are largely different, cortical representations of outcome valence signals are likely converted into a common currency to drive learning in similar ways, which might further explain the comparable behavioral measures and EEG signals across the two contexts.

Brainstem functional connectivity dynamics

To further understand how our temporally resolved brainstem clusters are coupled with other brain areas and whether the strength of that coupling is associated with behavioral choices, we performed additional psychophysiological interaction analyses. We used the brainstem clusters identified in each time window as seed regions and the participants' stay/switch response in the next trial as psychological regressor. We designed this analysis to test the extent to which brainstem outcome representations could explain downstream choice behavior, in accordance with reinforcement learning theory.[53] In our analysis, a positive correlation indicates stronger coupling when the same stimulus is chosen in the next trial (stay), and a negative correlation indicates stronger coupling when the opposite stimulus is chosen in the next trial (switch).

During reward learning, in the first time window, we found a negative coupling between the BOLD activity in the LC cluster and the anterior cingulate gyrus, consistent with the role of this region in error monitor and switching behavior.[54,55] We found a positive coupling between the SN/VTA cluster in the first time window and inferior frontal gyrus. In the second time window, we found a positive coupling between the VTA cluster and brain areas typically implicated in reward processing (e.g., ventral striatum, globus pallidum, thalamus), consistent with their role in reinforcing the rewarded choice (Figure 5, see also Table S5).[41]

During punishment learning, we only found a negative coupling between RN clusters and cortical structures (e.g., frontal medial/orbital and temporal cortex) as well as subcortical structures (e.g., midbrain, ventral striatum, thalamus, amygdala, hippocampus). Similarly, the SN was negatively coupled primarily with cortical structures (e.g., frontal medial/orbital, temporal fusiform/inferior, and occipital cortex) (Figure 5, see also Table S5). These findings corroborate the interplay between the brainstem and other subcortical and cortical regions known to play a role in cognitive control by enhancing avoidance (or switch) behavior following negative outcomes,[30,56] and they further support the involvement of our identified brainstem clusters in reinforcement learning.

Although we showed above that choice behavior was modulated by the strength of coupling between the brainstem clusters and other brain areas, we also tested whether the BOLD activity obtained directly from individual brainstem clusters predicted stay/switch behavior (see STAR Methods). We found that BOLD estimates in the VTA cluster during omission of reward (i.e., negative outcomes in the second time window of the reward context) predicted stay/switch behavior ($\beta = -0.00033$, $p = 0.029$), such that lower BOLD VTA activity for negative outcomes was associated with higher probability of repeating the choice (stay) in the next reward trial. Across subjects, higher VTA activity in response to negative outcomes was positively correlated with overall accuracy in the reward context ($r_{bend} = 0.45$, $p = 0.0098$). These additional analyses further support the role of the VTA in modulating reward-based learning/behavior. We did not find any significant relations between the other individual brainstem nuclei and stay/switch responses, which suggests that, more generally, choice behavior does not depend solely on individual brainstem nuclei but most likely on the strength of their coupling with other brain areas, as we have demonstrated previously.[45]

Oscillatory activity analysis of outcome valence

While our work focuses primarily on the brainstem pathways associated with transient evoked EEG responses, previous EEG studies have also investigated the role of ongoing oscillatory modes in encoding unexpected positive and negative outcomes.[57–60] To test how these oscillatory phenomena, in particular theta and high-beta power, might manifest in our own data, we performed a separate set of analyses (see STAR Methods).

Specifically, we found that EEG theta power was higher for negative compared to positive outcomes ($\beta = -21$, $p < 0.001$). In contrast, high-beta power was, on average, higher for positive

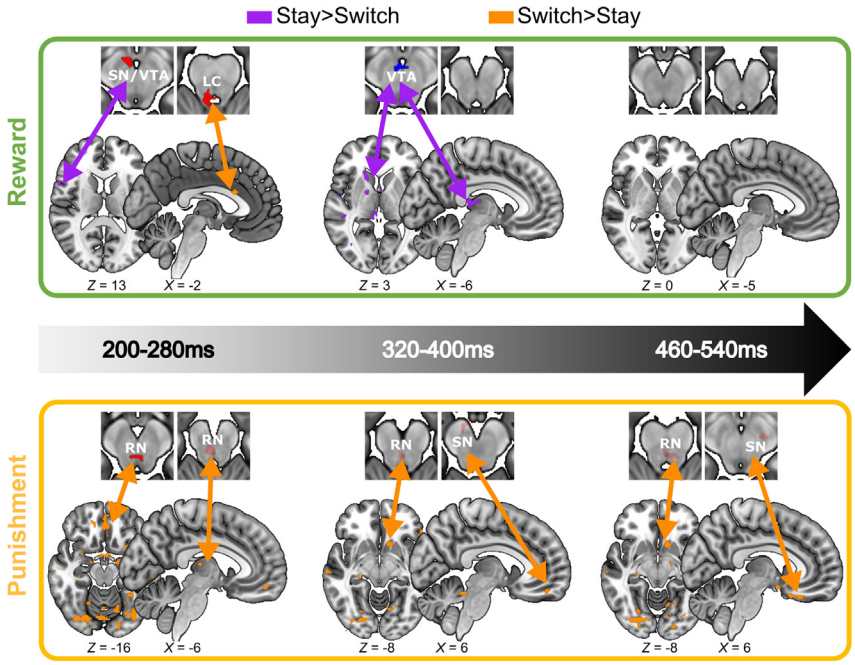


Figure 5. Brainstem functional connectivity and modulation of switch/stay behavior

Coupling between the brainstem clusters identified in the EEG-informed fMRI analysis in each time window and other brain networks (all *Z* scores >2.3, cluster corrected), when participants (n = 28) repeated (purple) or switched (orange) their choice in the next trial, in the reward (upper panel) and punishment (bottom panel) contexts (see also Table S5). Below each depicted slice, *Z* is the z coordinate in the standard MNI brain. SN/VTA, substantia nigra/ventral tegmental area; LC, locus coeruleus; RN, raphe nucleus.

compared to negative outcomes, although this difference did not reach statistical significance ($\beta = 1.6$, $p = 0.25$). Overall, these effects were prominent throughout the feedback period, in both the reward and punishment contexts (Figures S5A and S5B), and they are largely consistent with previous studies.[57–60] To assess the brain networks associated with these two oscillation modes, we subsequently performed an EEG-informed fMRI analysis by using single-trial power estimates in the theta and high-beta bands as parametric regressors of BOLD activity for positive and negative outcomes in each context.

Trial-wise fluctuations in theta power were associated primarily with activations in the punishment context, where we observed greater responses to positive than negative outcomes in occipital and cingulate areas. We also found greater responses to negative compared to positive outcomes in the caudate (Figure S5C). Trial-wise fluctuations in the high-beta power were associated with higher activations for positive than negative outcomes in the vmPFC (as in Mas-Herrero et al.[61]) in the reward context and in a separate caudate cluster (compared to theta power) in the punishment context. We also found higher activations for negative than positive outcomes in frontal clusters only in the punishment context (Figure S5C). Overall, these findings suggest that these oscillatory phenomena manifest largely at the level of the human reward network and are likely playing an additional modulating role on learning and valuation.[57,62]

Finally, we ran an exploratory analysis in which we showed that individual estimates for positive responses in the vmPFC associated with high-beta band in the reward context were negatively correlated with positive activations in the later VTA cluster identified by our original EEG-fMRI analysis in the second time window ($r = -0.37$, $p = 0.046$; linear regression, $\beta = -0.0036$, $p = 0.022$; Figure S5D). This seems consistent with a role of the dopaminergic VTA neurons in modulating vmPFC activity, which in turn might increase the high-beta power associated with reward processing.[57]

DISCUSSION

Here we coupled EEG with simultaneously acquired fMRI to offer a full characterization of the spatiotemporal dynamics of outcome valence signals in different brainstem subregions (SN/VTA, LC, and RN) during reward- and punishment-based learning. We leveraged the high-temporal resolution of the EEG to build fMRI regressors, which identified latent brainstem activations of outcome valence not seen with a stand-alone fMRI analysis and assigned temporal order to those activations.

Animal electrophysiology studies have proposed that SN/VTA signals positive and negative outcomes similarly in reward and punishment learning, representing a single dimension of value.[7] However, there is contradictory evidence suggesting that reward and punishment are represented as two independent dimensions in SN/VTA.[6] Our data are consistent with the latter, whereby SN/VTA exhibited early responses to negative outcomes and a distinct VTA cluster responded later to positive outcomes during reward learning, whereas during punishment learning, SN was activated later and by negative outcomes only. Furthermore, LC and RN responded to negative outcomes during reward and punishment learning, respectively, further reinforcing the notion of distinct brainstem pathways of outcome valence. This work challenges theories of outcome evaluation by suggesting that brainstem pathways might encode positive outcomes associated with delivery of rewards and omission of punishments differently. Similarly, our results suggest that negative outcomes associated with omission of rewards and delivery of punishments involve different pathways.

Critically, these insights are enabled by our EEG-informed fMRI fusion approach, which assigned temporal order to the relevant pathways and further revealed latent brainstem activations not observed with stand-alone fMRI. This further highlights the importance of using endogenous variability in the electrophysiological signal to build continuous fMRI regressors to capture more fine-grained variations in the BOLD signal than conventional categorical regressors, even in deep and small subcortical areas. For example, using this fusion approach in the reward context revealed that the LC and VTA emerge in serial

fashion over temporally distinct time windows. Early LC engagement likely enhances arousal and redirection of attention toward negative outcomes,[1] whereas subsequent VTA activation suggests a more deliberate process consistent with dopaminergic reward prediction error signaling (greater responses to positive than negative outcomes) observed in other human studies and more consistently in electrophysiological non-human animal studies.[5–7,16,19,42]

Consistent with this framing, we showed that early LC and SN/VTA responses to negative outcomes were functionally coupled with regions of the cingulate and frontal cortex, respectively, known target sites for error/salience detection and response inhibition.[1,54,63–66] Similarly, the later VTA cluster was functionally coupled with basal ganglia structures (nucleus accumbens and globus pallidum) and thalamus when participants repeated the choice in the next trial, and its BOLD responses for negative outcomes correlated with stay/switch behavioral patterns and overall accuracy, supporting the idea that VTA broadcasts signals associated with reward value updating and action reinforcement.[21,53] Taken together, these results can extend the framework of two separate valence systems during reward-based learning[45] to the brainstem, whereby early automatic arousal and salience LC and SN/VTA responses to negative outcomes interact with later VTA responses to promote avoidance behavior and downregulate value information during learning.

In the punishment context, EEG-informed regressors revealed the relative timing of outcome valence activations. Our results implicate distinct RN and SN subdivisions in the signaling of negative outcomes such that early and transient activation of dorsal RN neurons could signal rapid arousal and salience, whereas more sustained activation of median RN could be involved in adjusting behavioral responses to recurrent negative outcomes.[25,27,67,68] Later activations in ventral/dorsal SN subdivisions might be associated with the processing of negative outcomes and/or aversive salience, respectively, which are critical for punishment-based learning.[7,21,23,69,70] Our data are broadly consistent with electrophysiological evidence that subpopulations of RN serotoninergic and SN dopaminergic neurons can be activated by unexpected negative outcomes during punishment learning.[3,6,7,23,24,71]

RN serotonergic neurons are known to regulate dopaminergic transmission in SN through direct projections.[31,72–76] Our connectivity analysis indeed showed that early RN responses interacted with SN to generate choice behavior away from negative outcomes. Thus, it is likely that an early serotonergic system initiates a fast arousal/avoidance response in the presence of negative outcomes, which in turn facilitates dopaminergic responses of a later aversive-salience-processing system to enhance avoidance behavior. Our connectivity analysis also suggests that both RN and SN broadcast extensively to other subcortical and cortical networks to shape avoidance responses, and that it is this coupling, rather than activity of individual brainstem nuclei, that drives choice behavior. Taken together these findings provide further support for the role of these brainstem subregions in punishment-based learning.[56,71]

There is general consensus that dopamine and serotonin can respond to outcome valence in reward and/or punishment contexts.[77,78] Electrophysiological studies have shown that SN/VTA dopaminergic responses are fast and transient, whereas RN serotoninergic responses are typically characterized by slower temporal dynamics and more sustained over time.[3,71] However, human studies using stand-alone fMRI studies have been unable to assess and further validate these temporal dynamics. Here, by fusing EEG-fMRI, we showed that different SN/VTA subdivisions indeed showed transient responses to negative or positive outcomes, whereas an RN cluster showed sustained activity in response to negative outcomes over time during punishment learning. Our results, therefore, extend the temporal dissociations reported in animal studies for the SN/VTA vs. RN to the human brain and provide further support for the involvement of dopamine in both reward and punishment learning[7,21] and serotonin specifically in punishment learning.[78,79]

Whereas the spatiotemporal dynamics of brainstem responses were notably different between reward and punishment contexts, other brain areas followed a broadly similar pattern of activations. Both in reward and punishment contexts, activations to negative outcomes were mostly observed earlier, whereas activations to positive outcomes were more robust in later time windows. This serial cascade supports a two-valence system for both reward- and punishment-based learning outside of the brainstem, whereby an early system (e.g., insula and prefrontal cortex) engages fast alertness and avoidance responses in the face of negative outcomes, while a later, more deliberate system (e.g., striatum and ventromedial prefrontal cortex) controls value updating and learning.[45,48,80] It is thus possible that the early system generates a fast alertness response in the presence of negative outcomes, and in parallel, it downregulates the late value-updating system to generate avoidance behavior and promote learning.

Our data suggest that while brainstem subregions might signal positive and negative outcomes differently in reward vs. punishment contexts, these outcome valence signals are eventually broadcasted to a domain-general network for updating value information and driving learning. For example, the orbitofrontal cortex is known to increase activity in response to delivery of rewards and omission of punishments,[81,82] suggesting similar signatures of outcome valence across reward and punishment contexts.[83–86] Thus, our findings help to extend the "common currency" framework,[87] whereby domain-specific brainstem representations of outcome valence are progressively transformed into domain-general value signals to drive learning.

Finally, we also explored the role of ongoing (persistent) oscillatory phenomena and found that EEG power in the theta and high-beta frequency bands was predictive of outcome valence, consistent with previous reports.[57–60] We also showed that these oscillatory phenomena manifest largely at the level of the human reward network, suggesting a possible modulating role on learning and valuation.[57,62] In line with this interpretation, we revealed a tentative relationship between high-beta oscillations in the vmPFC and later VTA responses to positive outcomes during reward learning. While this finding is consistent with the hypothesis that dopamine modulates high-beta power in reward processing,[57] our broader spatiotemporal representations also exhibited some disagreement with previous EEG-fMRI studies

on oscillation modes,[60,61] likely due to task differences (i.e., gambling tasks with highly surprising/unexpected outcomes vs. scarcer unexpected outcomes [mostly negative] in our instrumental learning task). Future neuroimaging studies will be required to reconcile these differences and directly address the putative relationship between brainstem structures and oscillatory phenomena during reward- and punishment-based learning.

Limitations of the study

At first sight, the early BOLD activations we found in the SN/VTA in response to reward omission might seem in disagreement with previous electrophysiology studies, which have shown that dopamine neurons in the SN/VTA fire in response to unexpected rewards and suppress their activity in response to unexpected omission of rewards.[6–8] This apparent discrepancy might be explained by the nature of our learning task, in which reward omissions (or negative outcomes in general) are the most unexpected/surprising type of outcome, as participants quickly learned the relevant stimulus-reward associations. Indeed, prior electrophysiology studies have also shown that dopaminergic midbrain neurons respond to unexpected negative outcomes, which has been interpreted as a salience signal.[7,20,21,23] Our study suggests that this salience signal (more typically found in response to punishments) might extend to omission of rewards as an early SN/VTA response, in line with the two-component reward response proposed by Schultz.[39]

An important caveat in our study is that that the timing of the cascade of constituent brainstem processes should be interpreted in relative (e.g., early vs. late) rather than absolute terms. This is because scalp representations informing our fMRI analysis are derived mainly from the cortical target sites of these subcortical neuronal representations. As such, the absolute timing of our EEG components will be delayed relative to their subcortical counterparts, as synaptic transmission times need to be factored in. This is consistent with electrophysiological studies showing that reward signals in cortical areas (e.g., PFC and ACC) occur at a slower timescale (around 300–400 ms) than in the SN/VTA.[19,88–91] Despite this limitation, however, the fusion of EEG-fMRI still allowed us to expose subcortical representations that would normally be difficult to detect with standalone fMRI, by exploiting the trial-by-trial covariation in their responses with that of their cortical target sites.

Conclusions

In conclusion, here we demonstrated that leveraging the endogenous variability in electrophysiologically derived measures of outcome valence, to inform the analysis of simultaneously acquired high-resolution fMRI data, can offer insights about the spatiotemporal dynamics of the brainstem pathways involved in learning. As such, our approach has the potential to open avenues for the study of subcortical brain dynamics, thereby bridging the gap between non-human animal and human studies. Critically, our findings can also help improve our understanding of how humans learn to make adaptive choices in reward and punishment contexts and ultimately how the mechanisms underlying those choices are affected in mental disorders associated with disruptions in learning and decision-making mechanisms (such as anxiety, depression, and drug addiction).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
- METHOD DETAILS
  - Probabilistic reversal-learning task
  - EEG data acquisition
  - MRI data acquisition
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral analysis
  - Reinforcement-learning algorithms
  - EEG data preprocessing
  - Single-trial EEG analysis
  - fMRI preprocessing
  - fMRI analysis
  - Standard fMRI analysis — GLM 1
  - EEG-informed fMRI analysis of activations in each component — GLM 2
  - EEG-informed fMRI analysis of activations across components — GLM 3
  - Regions of interest and corrections for multiple comparisons
  - Psychophysiological interaction analysis
  - Associations between individual brainstem nuclei activity and switch/stay behavior
  - Time-frequency decomposition analysis
  - EEG power-informed fMRI analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113589.

ACKNOWLEDGMENTS

This work was supported by the European Research Council (ERC, DyNeRfusion, 865003; M.G.P). We thank Frances Crabbe for optimizing MRI sequences and technical assistance during data collection.

AUTHOR CONTRIBUTIONS

Conceptualization, J.C. and M.G.P.; methodology, J.C. and M.P.G.; software, J.C. and M.P.G.; formal analysis, J.C.; investigation, J.C.; resources, M.P.G.; data curation, J.C.; writing – original draft, J.C.; writing – review and editing, J.C. and M.P.G.; visualization, J.C.; supervision, M.P.G.; project administration, M.P.G.; funding acquisition, M.P.G.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 23, 2023
Revised: October 5, 2023
Accepted: November 30, 2023

REFERENCES

1. Aston-Jones, G., and Cohen, J.D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. Annu. Rev. Neurosci. *28*, 403–450.
2. Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2010). Dopamine in motivational control: rewarding, aversive, and alerting. Neuron *68*, 815–834.
3. Cohen, J.Y., Amoroso, M.W., and Uchida, N. (2015). Serotonergic neurons signal reward and punishment on multiple timescales. Elife *4*, e06346.
4. Watabe-Uchida, M., Eshel, N., and Uchida, N. (2017). Neural circuitry of reward prediction error. Annu. Rev. Neurosci. *40*, 373–394.
5. Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature *482*, 85–88.
6. Fiorillo, C.D. (2013). Two dimensions of value: Dopamine neurons represent reward but not aversiveness. Science *341*, 546–549.
7. Matsumoto, M., and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature *459*, 837–841.
8. Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.
9. Glimcher, P.W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. USA *108*, 15647–15654.
10. Anlezark, G.M., Crow, T.J., and Greenway, A.P. (1973). Impaired learning and decreased cortical norepinephrine after bilateral locus coeruleus lesions. Science *181*, 682–684.
11. Bouret, S., and Richmond, B.J. (2015). Sensitivity of locus ceruleus neurons to reward value for goal-directed actions. J. Neurosci. *35*, 4005–4014.
12. Breton-Provencher, V., Drummond, G.T., Feng, J., Li, Y., and Sur, M. (2022). Spatiotemporal dynamics of noradrenaline during learned behaviour. Nature *606*, 732–738.
13. Berridge, C.W., and Waterhouse, B.D. (2003). The locus coeruleus-noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. Brain Res. Rev. *42*, 33–84.
14. Sara, S.J., and Bouret, S. (2012). Orienting and reorienting: The locus coeruleus mediates cognition through arousal. Neuron *76*, 130–141.
15. Usher, M., Cohen, J.D., Servan-Schreiber, D., Rajkowski, J., and Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. Science *283*, 549–554.
16. D'Ardenne, K., McClure, S.M., Nystrom, L.E., and Cohen, J.D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science *319*, 1264–1267.
17. Murphy, P.R., O'Connell, R.G., O'Sullivan, M., Robertson, I.H., and Balsters, J.H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. Hum. Brain Mapp. *35*, 4140–4154.
18. Payzan-LeNestour, E., Dunne, S., Bossaerts, P., and O'Doherty, J.P. (2013). The Neural representation of unexpected uncertainty during value-based decision making. Neuron *79*, 191–201.
19. Zaghloul, K.A., Blanco, J.A., Weidemann, C.T., McGill, K., Jaggi, J.L., Baltuch, G.H., and Kahana, M.J. (2009). Human substantia nigra neurons encode unexpected financial rewards. Science *323*, 1496–1499.
20. Fiorillo, C.D., Song, M.R., and Yun, S.R. (2013). Multiphasic temporal dynamics in responses of midbrain dopamine neurons to appetitive and aversive stimuli. J. Neurosci. *33*, 4710–4725.
21. Cox, J., and Witten, I.B. (2019). Striatal circuits for reward learning and decision-making. Nat. Rev. Neurosci. *20*, 482–494.
22. Ungless, M.A., Magill, P.J., and Bolam, J.P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. Science *303*, 2040–2042.
23. Menegas, W., Akiti, K., Amo, R., Uchida, N., and Watabe-Uchida, M. (2018). Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat. Neurosci. *21*, 1421–1430.
24. Verharen, J.P.H., Zhu, Y., and Lammel, S. (2020). Aversion hot spots in the dopamine system. Curr. Opin. Neurobiol. *64*, 46–52.
25. Kawai, H., Bouchekioua, Y., Nishitani, N., Niitani, K., Izumi, S., Morishita, H., Andoh, C., Nagai, Y., Koda, M., Hagiwara, M., et al. (2022). Median raphe serotonergic neurons projecting to the interpeduncular nucleus control preference and aversion. Nat. Commun. *13*, 7708.
26. Maswood, S., Barter, J.E., Watkins, L.R., and Maier, S.F. (1998). Exposure to inescapable but not escapable shock increases extracellular levels of 5-HT in the dorsal raphe nucleus of the rat. Brain Res. *783*, 115–120.
27. Paquelet, G.E., Carrion, K., Lacefield, C.O., Zhou, P., Hen, R., and Miller, B.R. (2022). Single-cell activity and network properties of dorsal raphe nucleus serotonin neurons during emotionally salient behaviors. Neuron *110*, 2664–2679.e8.
28. Schweimer, J.V., and Ungless, M.A. (2010). Phasic responses in dorsal raphe serotonin neurons to noxious stimuli. Neuroscience *171*, 1209–1215.
29. Azmitia, E.C., and Segal, M. (1978). An autoradiographic analysis of the differential ascending projections of the dorsal and median raphe nuclei in the rat. J. Comp. Neurol. *179*, 641–667.
30. Cools, R., Roberts, A.C., and Robbins, T.W. (2008). Serotoninergic regulation of emotional and behavioural control processes. Trends Cognit. Sci. *12*, 31–40.
31. Van der Kooy, D., and Hattori, T. (1980). Dorsal raphe cells with collateral projections to the caudate-putamen and substantia nigra: A fluorescent retrograde double labeling study in the rat. Brain Res. *186*, 1–7.
32. Cools, R., Robinson, O.J., and Sahakian, B. (2008). Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. Neuropsychopharmacology *33*, 2291–2299.
33. Crockett, M.J., Clark, L., and Robbins, T.W. (2009). Reconciling the role of serotonin in behavioral inhibition and aversion: Acute tryptophan depletion abolishes punishment-induced inhibition in humans. J. Neurosci. *29*, 11993–11999.
34. Evers, E.A.T., Cools, R., Clark, L., van der Veen, F.M., Jolles, J., Sahakian, B.J., and Robbins, T.W. (2005). Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. Neuropsychopharmacology *30*, 1138–1147.
35. Faulkner, P., and Deakin, J.F.W. (2014). The role of serotonin in reward, punishment and behavioural inhibition in humans: Insights from studies with acute tryptophan depletion. Neurosci. Biobehav. Rev. *46 Pt 3*, 365–378.
36. Robinson, O.J., Cools, R., and Sahakian, B.J. (2012). Tryptophan depletion disinhibits punishment but not reward prediction: implications for resilience. Psychopharmacology *219*, 599–605.
37. Seymour, B., Daw, N.D., Roiser, J.P., Dayan, P., and Dolan, R. (2012). Serotonin selectively modulates reward value in human decision-making. J. Neurosci. *32*, 5833–5842.
38. Tanaka, S.C., Shishida, K., Schweighofer, N., Okamoto, Y., Yamawaki, S., and Doya, K. (2009). Serotonin affects association of aversive outcomes to past actions. J. Neurosci. *29*, 15669–15674.
39. Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci. *17*, 183–195.

40. Chandler, D.J. (2016). Evidence for a specialized role of the locus coeruleus noradrenergic system in cortical circuitries and behavioral operations. Brain Res. *1641*, 197–206.
41. Haber, S.N., and Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. Neuropsychopharmacology *35*, 4–26.
42. Schultz, W. (2002). Getting formal with dopamine and reward. Neuron *36*, 241–263.
43. Zubair, M., Murris, S.R., Isa, K., Onoe, H., Koshimizu, Y., Kobayashi, K., Vanduffel, W., and Isa, T. (2021). Divergent whole brain projections from the ventral midbrain in macaques. Cerebr. Cortex *31*, 2913–2931.
44. Bentivoglio, M., and Morelli, M. (2005). The organization and circuits of mesencephalic dopaminergic neurons and the distribution of dopamine receptors in the brain. In Handbook of Chemical Neuroanatomy Dopamine, S.B. Dunnett, M. Bentivoglio, A. Björklund, and T. Hökfelt, eds. (Elsevier), pp. 1–107.
45. Fouragnan, E., Retzler, C., Mullinger, K., and Philiastides, M.G. (2015). Two spatiotemporally distinct value systems shape reward-based learning in the human brain. Nat. Commun. *6*, 8107.
46. Pisauro, M.A., Fouragnan, E., Retzler, C., and Philiastides, M.G. (2017). Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. Nat. Commun. *8*, 15808.
47. Chen, Y., Chaudhary, S., and Li, C.-S.R. (2022). Shared and distinct neural activity during anticipation and outcome of win and loss: A meta-analysis of the monetary incentive delay task. Neuroimage *264*, 119764.
48. Fouragnan, E., Retzler, C., and Philiastides, M.G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. Hum. Brain Mapp. *39*, 2887–2906.
49. Liu, X., Hairston, J., Schrier, M., and Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. Neurosci. Biobehav. Rev. *35*, 1219–1236.
50. Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., and Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. Hum. Brain Mapp. *39*, 3398–3418.
51. Rutledge, R.B., Dean, M., Caplin, A., and Glimcher, P.W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. J. Neurosci. *30*, 13525–13536.
52. Clithero, J.A., and Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. Soc. Cognit. Affect Neurosci. *9*, 1289–1302.
53. Maia, T.V., and Frank, M.J. (2011). From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. *14*, 154–162.
54. Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747–749.
55. Kolling, N., Behrens, T., Wittmann, M.K., and Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. Curr. Opin. Neurobiol. *37*, 36–43.
56. Hormigo, S., Vega-Flores, G., and Castro-Alamancos, M.A. (2016). Basal ganglia output controls active avoidance behavior. J. Neurosci. *36*, 10274–10284.
57. Marco-Pallarés, J., Münte, T.F., and Rodríguez-Fornells, A. (2015). The role of high-frequency oscillatory activity in reward processing and learning. Neurosci. Biobehav. Rev. *49*, 1–7.
58. HajiHosseini, A., Rodríguez-Fornells, A., and Marco-Pallarés, J. (2012). The role of beta-gamma oscillations in unexpected rewards processing. Neuroimage *60*, 1678–1685.
59. Cohen, M.X., Elger, C.E., and Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. Neuroimage *35*, 968–978.
60. Andreou, C., Frielinghaus, H., Rauh, J., Mußmann, M., Vauth, S., Braun, P., Leicht, G., and Mulert, C. (2017). Theta and high-beta networks for feedback processing: a simultaneous EEG-fMRI study in healthy male subjects. Transl. Psychiatry *7*, e1016.
61. Mas-Herrero, E., Ripollés, P., HajiHosseini, A., Rodríguez-Fornells, A., and Marco-Pallarés, J. (2015). Beta oscillations and reward processing: Coupling oscillatory activity and hemodynamic responses. Neuroimage *119*, 13–19.
62. Cohen, M.X., Wilmes, K., and Vijver, I.v.d. (2011). Cortical electrophysiological network dynamics of feedback learning. Trends Cognit. Sci. *15*, 558–566.
63. Gompf, H.S., Mathai, C., Fuller, P.M., Wood, D.A., Pedersen, N.P., Saper, C.B., and Lu, J. (2010). Locus ceruleus and anterior cingulate cortex sustain wakefulness in a novel environment. J. Neurosci. *30*, 14543–14551.
64. Grueschow, M., Kleim, B., and Ruff, C.C. (2020). Role of the locus coeruleus arousal system in cognitive control. J. Neuroendocrinol. *32*, e12890.
65. Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., and Owen, A.M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. Neuroimage *50*, 1313–1319.
66. Aron, A.R., Fletcher, P.C., Bullmore, E.T., Sahakian, B.J., and Robbins, T.W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. Nat. Neurosci. *6*, 115–116.
67. Deakin, J.F., and Graeff, F.G. (1991). 5-HT and mechanisms of defence. J. Psychopharmacol. *5*, 305–315.
68. Cho, J.R., Treweek, J.B., Robinson, J.E., Xiao, C., Bremner, L.R., Greenbaum, A., and Gradinaru, V. (2017). Dorsal raphe dopamine neurons modulate arousal and promote wakefulness by salient stimuli. Neuron *94*, 1205–1219.e8.
69. Lerner, T.N., Shilyansky, C., Davidson, T.J., Evans, K.E., Beier, K.T., Zalocusky, K.A., Crow, A.K., Malenka, R.C., Luo, L., Tomer, R., and Deisseroth, K. (2015). Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell *162*, 635–647.
70. Zhang, Y., Larcher, K.M.-H., Misic, B., and Dagher, A. (2017). Anatomical and functional organization of the human substantia nigra and its connections. Elife *6*, e26653.
71. Matias, S., Lottem, E., Dugué, G.P., and Mainen, Z.F. (2017). Activity patterns of serotonin neurons underlying cognitive flexibility. Elife *6*, e20552.
72. Alex, K.D., and Pehek, E.A. (2007). Pharmacologic mechanisms of serotonergic regulation of dopamine neurotransmission. Pharmacol. Ther. *113*, 296–320.
73. Di Giovanni, G., Esposito, E., and Di Matteo, V. (2010). Role of serotonin in central dopamine dysfunction. CNS Neurosci. Ther. *16*, 179–194.
74. Dray, A., Gonye, T.J., Oakley, N.R., and Tanner, T. (1976). Evidence for the existence of a raphe projection to the substantia nigra in rat. Brain Res. *113*, 45–57.
75. Fibiger, H.C., and Miller, J.J. (1977). An anatomical and electrophysiological investigation of the serotonergic projection from the dorsal raphe nucleus to the substantia nigra in the rat. Neuroscience *2*, 975–987.
76. Lavoie, B., and Parent, A. (1990). Immunohistochemical study of the serotoninergic innervation of the basal ganglia in the squirrel monkey. J. Comp. Neurol. *299*, 1–16.
77. Cools, R., Nakamura, K., and Daw, N.D. (2011). Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology *36*, 98–113.
78. Boureau, Y.-L., and Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. Neuropsychopharmacology *36*, 74–97.
79. Daw, N.D., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. Neural Network. *15*, 603–616.
80. Fouragnan, E., Queirazza, F., Retzler, C., Mullinger, K.J., and Philiastides, M.G. (2017). Spatiotemporal neural characterization of prediction error valence and surprise during reward learning in humans. Sci. Rep. *7*, 4762.

81. Kim, H., Shimojo, S., and O'Doherty, J.P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. *4*, e233.

82. Mowrer, S.M., Jahn, A.A., Abduljalil, A., and Cunningham, W.A. (2011). The value of success: Acquiring gains, avoiding Losses, and simply being successful. PLoS One *6*, e25307.

83. Bartra, O., McGuire, J.T., and Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage *76*, 412–427.

84. Kahnt, T., Park, S.Q., Haynes, J.-D., and Tobler, P.N. (2014). Disentangling neural representations of value and salience in the human brain. Proc. Natl. Acad. Sci. USA *111*, 5000–5005.

85. Leknes, S., Lee, M., Berna, C., Andersson, J., and Tracey, I. (2011). Relief as a reward: Hedonic and neural responses to safety from pain. PLoS One *6*, e17870.

86. Seymour, B., O'Doherty, J.P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., and Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat. Neurosci. *8*, 1234–1240.

87. Levy, D.J., and Glimcher, P.W. (2012). The root of all value: a neural common currency for choice. Curr. Opin. Neurobiol. *22*, 1027–1038.

88. Amiez, C., Joseph, J.-P., and Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. Eur. J. Neurosci. *21*, 3447–3452.

89. Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. Nat. Neurosci. *10*, 647–656.

90. Schultz, W., Tremblay, L., and Hollerman, J.R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. Cerebr. Cortex *10*, 272–284.

91. Kobayashi, S., Nomoto, K., Watanabe, M., Hikosaka, O., Schultz, W., and Sakagami, M. (2006). Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. Neuron *51*, 861–870.

92. Smith, S., Jenkinson, M., Beckmann, C., Miller, K., and Woolrich, M. (2007). Meaningful design and contrast estimability in FMRI. Neuroimage *34*, 127–136.

93. Mullinger, K.J., Yan, W.X., and Bowtell, R. (2011). Reducing the gradient artefact in simultaneous EEG-fMRI by adjusting the subject's axial position. Neuroimage *54*, 1942–1950.

94. Parra, L.C., Spence, C.D., Gerson, A.D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. Neuroimage *28*, 326–341.

95. Philiastides, M.G., Tu, T., and Sajda, P. (2021). Inferring macroscale brain dynamics via fusion of simultaneous EEG-fMRI. Annu. Rev. Neurosci. *44*, 315–334.

96. Sajda, P., Philiastides, M.G., and Parra, L.C. (2009). Single-trial analysis of neuroimaging data: inferring neural networks underlying perceptual decision-making in the human brain. IEEE Rev. Biomed. Eng. *2*, 97–109.

97. Sajda, P., Gerson, A.D., Philiastides, M.G., and Parra, L.C. (2007). Single-trial analysis of EEG during rapid visual discrimination: Enabling cortically coupled computer vision. In Toward Brain-Computer Interfacing (The MIT Press), pp. 423–444.

98. Duda, R.O., Hart, P.E., and Stork, D.G. (2000). Pattern Classification, 2nd ed. (Wiley).

99. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimisation for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825–841.

100. Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., and Smith, S.M. (2012). FSL. Neuroimage *62*, 782–790.

101. Limbrick-Oldfield, E.H., Brooks, J.C.W., Wise, R.J.S., Padormo, F., Hajnal, J.V., Beckmann, C.F., and Ungless, M.A. (2012). Identification and characterisation of midbrain nuclei using optimised functional magnetic resonance imaging. Neuroimage *59*, 1230–1238.

102. Andersson, J.L.R., Jenkinson, M., and Stephen, S. (2007). Non-linear Registration Aka Spatial Normalisation (FMRIB Centre).

103. Woolrich, M.W., Ripley, B.D., Brady, M., and Smith, S.M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. Neuroimage *14*, 1370–1386.

104. Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Jenkinson, M., and Smith, S.M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage *21*, 1732–1747.

105. Bianciardi, M., Strong, C., Toschi, N., Edlow, B.L., Fischl, B., Brown, E.N., Rosen, B.R., and Wald, L.L. (2018). A probabilistic template of human mesopontine tegmental nuclei from in vivo 7T MRI. Neuroimage *170*, 222–230.

106. Tona, K.-D., Keuken, M.C., de Rover, M., Lakke, E., Forstmann, B.U., Nieuwenhuis, S., and van Osch, M.J.P. (2017). In vivo visualization of the locus coeruleus in humans: quantifying the test-retest reliability. Brain Struct. Funct. *222*, 4203–4217.

107. Cox, R.W., Chen, G., Glen, D.R., Reynolds, R.C., and Taylor, P.A. (2017). fMRI clustering and false-positive rates. Proc. Natl. Acad. Sci. USA *114*, E3370–E3371.

108. Mitra, P.P., and Pesaran, B. (1999). Analysis of dynamic brain imaging data. Biophys. J. *76*, 691–708.

109. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.

110. Cohen, M.X. (2014). Analyzing Neural Time Series Data: Theory and Practice (The MIT Press).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Human Subjects	N/A	N/A
Software and algorithms		
Custom computer code	This manuscript	https://doi.org/10.5281/zenodo.10203476
FSL	University of Oxford	https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
AFNI 3dClustSim	National Institutes of Health	https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dClustSim.html
MRIcroGL	Chris Rorden	https://www.nitrc.org/projects/mricrogl
EEGLAB v14.1.2.	Delorme and Makeig	https://sccn.ucsd.edu/eeglab/index.php
FieldTrip	Donders Institute for Brain, Cognition and Behavior	https://www.fieldtriptoolbox.org/
Presentation	Neurobehavioral Systems	https://www.neurobs.com/
Brain Vision Recorder	Brain Products	https://brainvision.com/products/recorder/
MATLAB	MathWorks	https://uk.mathworks.com/products.html?s_tid=gn_ps
Other		
MR-compatible EEG amplifier system	BrainAmps MR-Plus, Brain Products	N/A
Siemens 3-Tesla TIM Trio MRI scanner	Siemens	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Joana Carvalheiro (joana.carvalheiro@glasgow.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Behavioral and neuroimaging data is available upon request to the lead contact
- Original code for the linear discriminant analysis has been deposited at Zenodo and is publicly available. DOIs are listed in the key resources table.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Thirty-one subjects participated in the experiment. Three were removed from the analysis for average performance below chance in reward or punishment blocks (using *myBinomTest* function in MATLAB). The remaining 28 subjects (18 female and 10 male), aged between 18 and 36 years (mean = 25 years, s.d. ± 5.2), were included in all subsequent analyses. All were right handed, had normal or corrected-to-normal vision and reported no history of psychiatric, neurological or major medical problems, and were free of psychoactive medications at the time of the study. This study used a mixed sample of male and female participants as we had no *a priori* reason to predict differences between sexes. The study was approved by the College of Science and Engineering Ethics Committee at the University of Glasgow (300210062) and informed consent was obtained from all participants.

METHOD DETAILS

Probabilistic reversal-learning task

The experiment consisted of 4 blocks of 80 trials each (320 trials in total). Two blocks involved rewards and the other two involved punishments. Reward and punishment blocks were administered in alternated order and counterbalanced across participants. Blocks were separated by a break. At the beginning of each block, subjects were shown a screen with two abstract symbols (the same two abstract symbols were used in all blocks). Subjects were told that their goal was to seek the symbol with the highest reward probability in reward blocks, and to avoid the symbol with the highest punishment probability in punishment blocks. They were also informed that in the course of each block, the highest-probability symbol might shift to the other symbol and that they would have to adjust their choices accordingly. In reward blocks, they could win 1 point or nothing; in punishment blocks, they could lose 1 point or nothing. Subjects were told that they would receive a fixed payment for participation (£20), and that losses (points converted into £'s) would be subtracted from that amount and wins added to that amount. They were also told that this additional amount would vary between £10–50 and would be based on the outcome of a random subset of trials selected at the end of the experiment. No further details regarding the mapping between earned points and the final monetary payoff were given to the subjects. Every participant was paid the same total amount (£40).

At the start of each block, a message highlighted whether it was a reward or a punishment block. Each trial began with the presentation of a central fixation cross for a random delay in the range 1–4 s (mean delay 2.5 s). To ensure alertness during the experiment and minimize saccades, subjects were instructed to focus on the central fixation. The two symbols were then placed to the left and to the right of the fixation cross for 1.25 s. During this time, subjects had to choose one of the symbols by pressing the left or right button on a response box using their right index or middle finger, respectively. Next, the choice outcome was presented after a second random delay in the range 1–4 s (mean delay 2.5 s). Trials, in which subjects did not respond within the 1.25 s of the stimulus presentation, were followed by a 'too slow' message and were excluded from further analysis. To increase detection power and estimation efficiency in the fMRI analysis, the sequence of these events and the timing of the two delay periods were optimised using standard efficiency equations.[92] Positive and negative outcomes were provided by displaying different arrows in the center of the screen for 750 ms. Specifically, in reward blocks, we used upward and neutral arrows to respectively provide positive and negative feedback, and in punishment blocks we used neutral and downwards arrows to respectively provide positive and negative feedback. All arrows were normalised for perceptual load, by using the same pixel count and overall structure.

At any one point in the course of the experiment, one of the two symbols was associated with a 'high' reward/punishment probability of 0.7 (that is, good/bad symbol) and the other symbol had a reward/punishment probability of 0.3 (that is, bad/good symbol). Participants were not informed about the exact probabilities assigned to each symbol and they were told to learn to choose the good symbol and avoid the bad symbol through trial and error by taking into account the outcome of their choices in every trial. Reversals were introduced by changing the contingencies of the symbols, that is the symbol with the highest reward/punishment probability symbol was assigned the lowest reward/punishment probability, and vice-versa for the other symbol. Reversals were programmed to occur every 20 ± 2 trials, and they occurred three times per block. To familiarise participants with the task, participants were asked to complete an online version of the task with one reward and one punishment block, 60 trials each, on the day before the experiment. An additional practice was completed before they entering the scanner, to ensure that participants understood the probabilistic nature of the task.

EEG data acquisition

EEG was collected simultaneously with the fMRI data using an MR-compatible EEG amplifier system (BrainAmps MR-Plus, Brain Products) and recorded using Brain Vision Recorder (Brain Products) with a 5-kHz sampling rate. Data were filtered online with a hardware band-pass filter of 0.016–250 Hz. The EEG cap included 63 Ag/AgCl scalp electrodes which were localized according to the international 10–20 system, and an electrocardiogram electrode, which was positioned along the paravertebral line. All electrodes had in-line 10 kΩ surface-mount resistors to ensure subject safety, which was further guaranteed by bundling and twisting all leads for their entire length. We lowered the input impedance for each electrode to <50 kΩ. The acquisition of EEG and MRI data were synchronized (Syncbox, Brain Products) and MR-scanner triggers were recorded separately for the subsequent offline removal of MR gradient artifacts. Scanner pulses were lengthened to 50 μs via an in-house built pulse stretcher. Experimental event codes and participants' responses were synchronized, and recorded simultaneously, with the EEG data through the Brain Vision Recorder software. Subjects were positioned inside the scanner by ensuring that electrodes Fp1 and Fp2 were aligned with the isocenter of the MR scanner.[93] The ribbon cable connecting to the EEG amplifiers at the back of the bore was secured to a cantilever beam to minimize scanner vibration artifacts. The helium pumps of the MRI scanner was switched off during acquisition, to minimise EEG artifacts in the high-frequency ranges.

MRI data acquisition

A Siemens 3-Tesla TIM Trio MRI scanner with a 12-channel head coil was employed for the fMRI acquisition. Because the focus of our study was the brainstem, we acquired T2*-weighted echo planar images (EPI) with coverage limited to the midbrain and upper portion of the pons while subjects were performing the task. This coverage also included part of the prefrontal cortex, striatum and globus pallidus, thalamus, insula, and amygdala (among other regions) (see Figure S2A). A total of 30 slices were acquired with an

interleaved-ascending order for each T2*-weighted EPI volume, with an isotropic resolution of 2 mm. Other imaging parameters included the following: TR, 2000 ms; TE, 30 ms; flip angle, 77°; field of view, 216 × 216mm; matrix, 108 × 108 mm. A whole-brain EPI with similar parameters was acquired for registration purposes. Using a 32-channel coil, we acquired a partial coverage high-resolution T2-weighted structural scan (T2-weighted 3D SPACE, isotropic voxel size, 0.75 mm) and a high-resolution T1-weighted structural scan (isotropic voxel size, 1 mm), which were used to optimise co-registration to brainstem structures (see below).

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analysis

To compare participants' performance and response times between reward and punishment contexts, we averaged the percentage of "correct" choices (i.e., choice of the stimulus with high-probability of reward and low-probability of punishment) and response times for each subject (n = 28 subjects) and performed paired t-tests (normality assumption confirmed with Shapiro-Wilk, $p > 0.14$). To test whether switch choices, response times and update of expected value (see section below for a description of the reinforcement-learning model used to estimate expected values) were affected differently by positive and negative outcomes, and whether this effect differed between reward and punishment contexts, we conducted repeated-measures ANOVAs with outcome valence (positive and negative) and context (reward and punishment) as within-subject factors to test the outcome valence × context interaction. For this, we a) extracted the percentage of switch choices away from the symbol that led to positive and negative outcomes (i.e., whether participants chose the opposite symbol in the next trial), b) averaged response times in trials that followed positive and negative outcomes, and c) calculated the absolute difference between the expected values on trials that followed positive and negative outcomes and the current trial (i.e., the extent to which participants updated the value of their choice after receiving positive or negative outcomes).

Reinforcement-learning algorithms

We used a standard reinforcement-learning algorithm to estimate trial-by-trial choice values using each subject's behavioral responses.

Specifically, each pair-stimulus value, or *Q*-value, was initialized to zero, and for each trial, t, within that pair of stimuli, the value of the chosen stimulus (say A was chosen) was updated according to:

$$Q_A\ (t\ +\ 1) = Q_A(t) + \alpha * \delta(t), \quad \text{(Equation 1)}$$

where δ was the prediction error:

$$\delta(t) = r(t)\ -\ Q_A\ (t), \quad \text{(Equation 2)}$$

where $r(t)$ was 1 for wins, 0 for neutral outcomes, and -1 for losses. The learning rate, α, was given by:

$$\alpha = \begin{cases} \alpha^{+}, if\ \delta(t) > 0 \\ \alpha^{-}, if\ \delta(t) < 0 \end{cases}, \quad \text{(Equation 3)}$$

where α^{+} and α^{-} were the learning rates for positive and negative prediction errors, respectively.

The probability of choosing one stimulus over another (say A over B) was given by the *softmax* equation:

$$P_A(t) = \frac{e^{[Q_A(t) * \beta]}}{e^{[Q_A(t) * \beta]} + e^{[Q_B(t) * \beta]}}, \quad \text{(Equation 4)}$$

where the β parameter, or inverse temperature, reflects the noise in choice selection.

Model fitting involved estimating the values of the free parameters (α^{+}, α^{-}, and β) that best accounted for the respective trial-by-trial choices in each context, using maximum likelihood estimation and a constrained non-linear optimization procedure (as implemented in *fmincon* in MATLAB) separately for each subject. To assess the goodness of fit, we compared the choice probabilities predicted by the reinforcement-learning model using the *softmax* procedure to subjects' behavioral choices by binning *P* (Equation 4) into 10 bins (bin size of 0.1) and calculating for each bin the fraction of trials in which subjects actually chose one of the stimulus (Figure S1A). Trial-by-trial *Q*-values obtained in Equation 1 were extracted for the behavioral analysis described above.

EEG data preprocessing

EEG data was preprocessed offline using MATLAB (Mathworks). EEG signals recorded inside an MR scanner are contaminated with gradient and ballistocardiogram (BCG) artifacts due to magnetic induction on the EEG leads. We first removed the gradient artifacts. To correct for gradient artifacts, we built artifact templates from sets of 80 consecutive functional volumes centered on each volume of interest, and subtracted these from the EEG signal. This process was repeated for as many times as there were functional volumes in our datasets. We subsequently applied a 10-ms median filter to remove any residual spike artifacts. Next, we band-pass filtered the data by applying a 0.5-Hz Butterworth high-pass filter to remove slow direct current drifts and a 40 Hz Butterworth low-pass filter to remove higher frequency noise. All data were downsampled to 1000Hz.

To remove eye blinks, we asked participants to perform an eye movement calibration task inside the scanner before the main experiment during which they were instructed to blink repeatedly several times while a central fixation cross was displayed in the center of the computer screen. We recorded the timing of these events and used principal component analysis to identify linear components associated with eye-blinks, which were subsequently removed from the broadband EEG data collected during the task.[45,94]

To deal with BCG artifacts in our data, we adopted a conservative approach which removed only a small number of BCG components (between 1 and 4) using principal component analysis in order to avoid removing physiologically relevant signals with more aggressive removal pipelines. Instead, we relied on our multivariate discriminant analysis, which was specifically designed to tolerate small residual artifacts that are not systematically aligned to experimental events of interest. Specifically, by integrating signals across sensors our discrimination procedure finds a low dimensional space that is orthogonal to any artifact residuals.[95] BCG principal components were extracted after the data were low-pass filtered at 4 Hz to extract the signal within the frequency range where most of the BCG power residues. Then, subject-specific principal components were determined based on the number of components that achieved the best discrimination (average number of components across subjects: 3.4) on a single-trial classifier for reward-positive vs. punishment-negative outcomes within the interval 100–600 ms post-outcome onset. The sensor weightings corresponding to the derived subject-specific principal components were projected onto the broadband (original) data and subtracted out. Comparison of the spectral profile of our EEG-fMRI data after artifact removal, with that of an EEG stand-alone dataset collected in our lab during administration of the same task (where no gradient or BCG artifacts are present), confirmed that the EEG-fMRI data was of sufficient quality to proceed with all the analyses below (see Figures S1C and S1D).

Single-trial EEG analysis

We applied a single-trial multivariate discriminant analysis, combined with a sliding window approach[94,96,97] to discriminate between positive and negative outcome trials in the EEG data locked to the time of choice outcome in reward and punishment contexts. Data were concatenated across the two blocks in each context, resulting in one analysis for reward and one for punishment. This method estimates, for predefined time windows, an optimal linear combination of EEG sensor weights (i.e., a spatial filter) which, applied to the multichannel EEG data [$x(t)$], yields a one-dimensional projection [i.e., a discriminant component $y(t)$] that discriminates between positive and negative outcomes:

$$y(t) = w^T x(t) = \sum_{i=1}^{D} w_i x_i(t), \quad \text{(Equation 5)}$$

where D represents the number of channels, indexed by i, and T indicates the transpose of the matrix. We applied this method to identify w for short (60 ms) overlapping time windows centered at 10 ms increments, between −100 and 850 ms relative to the onset of the choice outcome. This procedure was repeated for each subject, time window, and reward ad punishment contexts separately. When applied to an individual trial, spatial filters w can produce a measurement of the discriminant component amplitude for that trial. When separating between positive and negative outcomes, the discriminator was designed to map the component amplitudes for one outcome type to positive values and those of the other outcome type to negative values. Here, we mapped the positive outcome trials to positive values and the negative outcome trials to negative values, but note that this mapping is arbitrary.

We used this approach to identify all time windows yielding successful discrimination performance in the outcome period and used the resultant single-trial component amplitudes, $y(t)$, to construct parametrically modulated BOLD predictors for our fMRI analysis (see fMRI analysis section). To quantify the performance of the discriminator for each time window, we computed the area under a receiver operating characteristic (ROC) curve (i.e., the A_z value), using a leave-one-out-trial cross-validation procedure.[98] Specifically, for every iteration, we used N-1 trials to estimate a spatial filter w, which was then applied to the remaining trial to obtain out-of-sample discriminant component amplitudes y for positive- and negative-outcome trials and compute the A_z. The linearity of our model also allows computing scalp projections of the relevant discriminating components resulting from Equation 5 by estimating a forward model for each component:

$$a = \frac{Xy}{y^T y} \quad \text{(Equation 6)}$$

where the EEG data X and discriminating components y are now in a matrix and vector notation, respectively, for convenience (i.e., both X and y now contain a time dimension). Equation 6 describes the electrical coupling of the discriminating component y that explains most of the activity in X. Strong coupling indicates low attenuation of the component y and can be visualised as the intensity of vector a. The resulting scalp topographies were used to identify time windows with largely distinct spatial profiles, which is indicative of distinct neural networks.

fMRI preprocessing

We discarded the first 5 EPI volumes in each block before data processing and statistical analysis to allow for magnetisation equilibration, and the remaining 290 volumes were used for the statistical analyses. Prior to the preprocessing, data from the two blocks within each context (reward or punishment) were concatenated for consistency with the way the EEG data were analyzed and to maximise power in our subsequent EEG-informed fMRI analysis, which included the full range of trial-wise amplitude estimates of

the relevant EEG components. To concatenate the data, we first normalised the data intensity for each block. Then, we transformed an exemplar EPI in each block to an exemplar EPI from the first block in each context (using linear registration) and the resulting transforms were applied to all EPIs (to transform EPIs from different blocks to the same space). EPIs across the two blocks within each context were then concatenated, resulting in one dataset for the reward context and another dataset for the punishment context (580 volumes in each dataset). Head motion parameters were estimated for each block using the MCFLIRT tool[99] and then concatenated for each context.

Pre-processing of the MRI data was performed on each concatenated dataset using the FEAT tool of the FSL software (FMRIB Software Library)[100] and included slice-timing correction, high-pass filtering (>100 s), and spatial smoothing (with a Gaussian kernel of 3 mm full width at half maximum). Brain extraction of the structural and functional images was performed using the Brain Extraction tool (BET). Registration of EPI images to standard space (Montreal Neurological Institute, MNI) was optimised to improve the registration of data with a limited FOV,[101] by using a whole-brain EPI and high-resolution structural images (T2-and T1-weighted) across different steps (see Figures S2B and S2C). Specifically, an exemplar EPI image was first registered to the whole-brain EPI (that matched the functional data in terms of contrast and resolution), the whole-brain EPI was registered to the T2w structural image, and the T2w structural image was registered to the T1w structural image, using linear registration (FLIRT command in FSL[99]) in all steps. The three resulting transforms were concatenated into a single transform to avoid image degradation through multiple transforms, so that in the end we had a single transform (EPIs transformed into T1w space). To have EPIs in the MNI space, T1w images were transformed into MNI space using Non-linear Image Registration Tool (FNIRT command in FSL[102]) with a 10 mm warp resolution, and the resulting transform was combined with the transform from EPIs to T1-W space, and then applied to all EPIs.

fMRI analysis

Statistical analyses of functional data were performed using a multilevel approach within the framework of a GLM, as implemented in FSL (using the FEAT module[103]):

$$Y = X\beta + \epsilon = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N + \epsilon \quad \text{(Equation 7)}$$

where Y represents the timeseries (with T time samples) for a voxel and X is a $T \times N$ design matrix where the columns correspond to the different regressors included in the design convolved with a canonical hemodynamic response function. β is a $N \times 1$ column vector of regression coefficients (i.e., betas or parameter estimates) and ϵ a $T \times 1$ column vector of residual error terms.

A first-level analysis was performed to analyze each subject's individual contexts (with the two blocks in that context concatenated, as described above). For each subject, we performed a GLM analysis for the reward context and another GLM for the punishment context, using the same framework in both (see below). To combine data across subjects within each context, a group-level, mixed-model was used (FLAME 1), treating participants as a random effect.[104] A full list of significant peaks for each GLM and contrast can be found in the supplementary material.

Standard fMRI analysis — GLM 1

We first performed a conventional fMRI analysis aimed at identifying differences in brainstem responses to positive and negative outcomes within reward and punishment contexts. Specifically, we set one GLM for the reward context and other GLM for the punishment context, both locking at the time of outcome we included two boxcar regressors of interest with a duration of 750 ms – to match the duration of the outcome stimulus – for each regressor event: (1) an unmodulated regressor for positive outcomes (all event amplitudes set to 1), and (2) an unmodulated regressor for negative outcomes (all event amplitudes set to 1). In addition, we included an unmodulated regressor of no interest at the time of stimulus presentation (that is, choice phase), an unmodulated regressor for all missed trials, an unmodulated regressor for the concatenation point for the two blocks within the context, and six nuisance regressors, one for each of the head motion parameters (three rotations and three translations). Contrasts were performed at first-level between the two unmodulated regressors of interest for positive and negative outcomes (positive > negative and negative > positive) and then combined at group-level.

EEG-informed fMRI analysis of activations in each component — GLM 2

In this analysis, we exploited the EEG single-trial-variability in three discriminating components of outcome valence to build EEG-informed fMRI regressors. Specifically, we used the resulting trial-by-trial amplitude estimates of $y(t)$ (Equation 5) for each outcome type (positive and negative), averaged across each of the three time windows. We replaced the two unmodulated outcome regressors in GLM 1 with these parametric regressors, resulting in a total of six regressors of interest. Because negative outcomes were mapped to negative y-values in the single-trial EEG analysis (i.e., more negative y's are indicative of a stronger response to a negative outcome), we flipped the sign of the y-values in the negative outcome regressors to better facilitate the interpretation of our contrasts. We set one GLM for the reward context and other for the punishment context. In each GLM, first-level contrasts were performed between positive and negative outcomes regressors (positive > negative and negative > positive) within each component. The rest of the design was identical to GLM 1.

EEG-informed fMRI analysis of activations across components — GLM 3

Because GLMs 2 included the trial-by-trial amplitude estimates of for all the three components, it reports activations in one component that are not explained by variance in other components. However, it is possible that some brain areas show sustained and similar activation across components, and therefore it is not captured by the GLM 2. To capture these putative "common" activations, we set separate GLMs for each component (that is, a total of six GLMs — three for the reward context and three for the punishment context). The design and contrasts were identical to GLM 2.

Regions of interest and corrections for multiple comparisons

Our analyses focused on the substantia nigra-ventral tegmental area (SN/VTA) complex, raphe nucleus (RN) and locus coeruleus (LC), brainstem subregions which have been implicated in reward and/or punishment learning. To build masks for these brainstem subregions, we used the Brainstem Navigator Atlas.[105] The SN/VTA mask was built from adding probabilistic bilateral masks for SN and VTA thresholded at 0.35 (total voxels: 320, 2mm isotropic). The RN mask was built from adding probabilistic bilateral masks for the dorsal raphe nucleus, caudal-rostral linear raphe and median raphe (total of 151 voxels, 2mm isotropic). As some of these RN are very small (RN nuclei vary between 12 and 98 voxels, 2mm isotropic), we did not apply a threshold on top of the original masks (otherwise some of the nuclei would be missed). For the LC mask, we used probabilistic bilateral LC masks from the probabilistic atlas in Tona et al.,[106] as the LC masks in Brainstem Navigator Atlas, even unthresholded, would result only in 30 voxels, which did not reach the minimum size (128 voxels) to run simulations for multiple comparisons correction (see below).

To control for false-positive rates, we determined cluster extent thresholds that were corrected for multiple comparisons by using 3dClustSim[107] in each of our brainstem masks and whole-brain (to cover our entire field-of-view). This method runs 10,000 Monte Carlo simulations that take into account the brain search volume and the degree of smoothing of the data to compute a cluster-size threshold for a given voxel-wise p value threshold, which we set at 0.01 (Z score = 2.3) for all analyses, such that the probability of surviving the threshold is $p < 0.05$. Across 5000 permutations, the 3dClustSim simulations determined the cluster sizes of 5 voxels for SN-VTA, 4 voxels for RN and 5 voxels for LC, and 21 voxels for the whole brain. We therefore used these results to derive corrected thresholds for our statistical maps and to select the clusters that showed overlap with our masks and survived the corrected threshold for the respective mask. As some clusters could extend beyond the brainstem, we additionally applied a brainstem mask (Harvard-Oxford subcortical structural atlas in FSL) to include only voxels that fell within the brainstem (for all reported clusters, their peak coordinates were originally located in the brainstem).

Psychophysiological interaction analysis

We extracted time-series data from individual clusters from each brainstem subregion of the three outcome-locked windows of interest, which served as seed regions (that is, physiological regressor). To extract the time-series data from subject-specific clusters, we back-projected the identified clusters at the group level in standard space into each individual's EPI space by applying the inverse transformations estimated during registration (see fMRI preprocessing section). As psychological regressor, we used the stay/switch choice in the next trial (collapsed across positive and negative outcomes), to model the strength of association between BOLD activity in brainstem clusters and other brain areas at the time of the outcome with the choice in the next trial. When participants chose the same stimulus in the next trial, we set the psychological regressor amplitude to 1 (stay) and when they chose the other stimulus in the next trial, we set the psychological regressor amplitude to -1 (switch).

The PPI analyses thus included the following regressors during the outcome phase: (1) an unmodulated regressor for choice outcome (all event amplitudes set to 1), (2) the physiological regressor (time-series in the EEG-fMRI derived clusters), (3) the psychological regressor (stay/switch choice in the next trial) and (4) the interaction regressor (physiological × psychological). The rest of the design was identical to GLM 1/2/3. Correction for multiple comparisons was performed on the entire fMRI field-of-view using the procedure described above ($Z > 2.3$, minimum 21 voxels).

Associations between individual brainstem nuclei activity and switch/stay behavior

To test whether the BOLD activity in the individual brainstem nuclei identified in the EEG-fMRI analysis predicted stay/switch behavior, we performed single-trial GLMs for each block in the reward and punishment contexts (4 GLMS in total). Each GLM included separate regressors for each individual trial (resulting in 80 regressors per GLM) and six nuisance regressors (one for each of the head motion parameters). We then extracted the single-trial BOLD estimates from the clusters identified in the EEG-fMRI analysis, for either negative or positive outcomes, as predictors of binary choice in the next trial (stay was coded as 1 and switch as 0) in generalized linear-mixed effects models with binomial distribution (logit function). Subject ID was included as random effect.

Time-frequency decomposition analysis

We extracted time-frequency information at the single-trial level from a cluster of four EEG electrodes (CP1, CP2, P1 and P2), appearing consistently on the scalp topographies of all EEG components identified in our main task (Figure 2). Specifically, we computed time-frequency representations in the theta (3.5 - 8 Hz) and high-beta (25 - 36 Hz) range using the multitaper approach[108] as implemented in the Fieldtrip package.[109] with a sequence of orthogonal Slepian tapers, a sliding fixed window length of 400 ms, and frequency smoothing of ± 4 Hz. The multitaper method can be better suited for estimating frequency representations characterised by

low signal to-noise ratio, as is the case with oscillatory signals in the higher frequency range.[110] We applied this method on outcome-locked data in the range −500 ms before to 1200 ms after the outcome onset.

At each time point and frequency layer we computed percentage change in power relative to a 400 ms pre-stimulus baseline. Thus, we considered task-relevant changes in power with respect to the pre-stimulus baseline, rather than absolute power. The effects of outcome valence (positive vs. negative) and context (reward vs. punishment) on theta and high-beta power were tested using separate linear-mixed effect models (LMMs). Dependent variables for the two LMMs were single-trial theta and high-beta power (averaged over time); valence (positive outcomes coded as 1 and negative outcomes as 0) and context (reward coded as 1 and punishment as 0) were fixed-effect factors and subject ID was included as a random factor. The valence × context interaction was initially included in the models but subsequently removed, as it was not significant in either of the two analyses ($p > 0.15$).

EEG power-informed fMRI analysis

We also performed a separate EEG-informed fMRI analysis using the power estimates obtained in the analysis above, whereby we followed the same pipeline as in our main analysis. Specifically, we built four parametric regressors corresponding to the single-trial oscillatory power between 200 and 525 ms (to overlap with the time range of our original analysis) in the theta and high-beta bands for separate positive and negative outcomes (as confounders, we included categorical regressors for stimulus and outcome presentation, concatenation point of the data and motion regressors), within each context (separate GLMs for reward and punishment contexts). We included the power from the two frequency bands in the same GLM, as they were not highly correlated at the single-trial level ($0.08 <$ average $r < 0.14$) and would additionally allow us to report activations unique to each frequency band. These parametric regressors were demeaned by subtracting the mean (theta or high-beta) oscillatory power within each valence condition from the single-trial power estimates for the corresponding valence. Similarly to our main analysis, we tested the contrasts positive > negative, and vice-versa, within each context and applied the same cluster correction method/threshold when reporting the results.