JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

The human cerebellum encodes temporally sensitive reinforcement learning signals

https://doi.org/10.1523/JNEUROSCI.2313-25.2026

Received: 19 December 2025
Revised: 5 May 2026
Accepted: 22 May 2026

The human cerebellum encodes temporally sensitive reinforcement learning signals

Juliana E. Trach[1], Yiran Ou[1], & Samuel D. McDougle[1,2]

1 – Dept of Psychology, Yale University, New Haven, CT 06510

2 – Wu Tsai Institute, Yale University, New Haven, CT 06510

Correspondence:

Juliana E. Trach: juliana.trach@yale.edu

Keywords: cerebellum; human neuroscience; reinforcement learning; reward prediction error

Running title: Reinforcement learning signals in the human cerebellum

Main text figures: 5

Supplemental figures: 5

Abstract: 235 words

Introduction: 650 words

Discussion: 1495 words

Declaration of interests

The authors have no competing interests.

Acknowledgements

We would like to thank the ACT lab and the other Cog Neuro labs at Yale for helpful discussions about this project. These data were collected at BrainWorks at the WuTsai Institute at Yale University. Thank you to Roeland Hancock and Alex Forrence for technical support at BrainWorks and to Tess Levy, Sanghoon Kang, Laurent Caplette, Lily Behm, Diana Wei, Omri Raccah, and Sophie Allen for support with data collection. JET is supported by the NSF GRFP. SDM is supported by NIH grant number R01 NS132926.

Abstract

In addition to supervised motor learning, the cerebellum also supports nonmotor forms of learning, including reinforcement learning (RL). Recent studies in animal models have identified core RL signals related to reward processing, reward prediction, and prediction errors in specific regions in cerebellar cortex. However, the constraints on these signals remain poorly understood, particularly in humans. Here, we investigated cerebellar RL signals in a computationally-driven fMRI study. Human participants performed an RL task without low-level sensorimotor contingencies (N = 32, N female = 24). We observed robust RL signals related to reward processing and reward prediction errors in cognitive regions of the cerebellum. These signals were not explained by oculomotor or physiological confounds. By manipulating the delay between choices and reward outcomes, we discovered that cerebellar RL signals are temporally sensitive: robust when feedback was delivered shortly following choices, but undetectable at supra-second feedback delays. Similar delay effects were not found in other areas implicated in reward processing, including the ventral striatum and hippocampus. Further, reward prediction error activity in the cerebellum was related to behavioral performance when feedback was delivered promptly, but not when it was delayed. Connectivity analyses revealed that during RL feedback, cognitive areas of the cerebellum coactivate with a network that includes the medial and lateral prefrontal cortex and caudate nucleus. Together, these results highlight a temporally constrained contribution of the human cerebellum to a cognitive learning task.

Significance statement

Reinforcement learning (RL) - the shaping of behavior through reward feedback - is an essential cognitive capacity. Previous work has focused almost exclusively on cerebral circuits that support RL, however recent work in animal models also implicates the cerebellum. Despite growing interest in the "cognitive cerebellum," its contributions to human RL remain unclear. We show that regions spanning lobules Crus I/II respond to rewards and encode reward prediction errors in a temporally sensitive manner. Our results build directly on research in model organisms, highlighting functional parallels across species and the importance of crosstalk between human and animal researchers. Including the cerebellum as a node in RL networks creates a potential new target for intervention in cases of reward processing dysfunction, like addiction.

Introduction

A recent paradigm shift in cerebellar neuroscience recognizes that the cerebellum contributes to both motor control and cognition (Buckner, 2013; Diedrichsen et al., 2019; Diedrichsen & McDougle, 2026; Ito, 2008; Leiner et al., 1986; Schmahmann, 2019; Sokolov et al., 2017; Stoodley et al., 2012; Strick et al., 2009). This shift is motivated by converging lines of evidence: The cerebellum is implicated in cognitive neurodevelopmental disorders (Stoodley, 2016), it is activated by nonmotor tasks (King et al., 2019; Van Overwalle et al., 2014, 2020), damage produces cognitive deficits (Butcher et al., 2017; Drepper et al., 1999; McDougle et al., 2022; Nicholas et al., 2023; Schmahmann & Sherman, 1998; Timmann et al., 2010), and it has bidirectional connections with nonmotor areas of the cerebrum (Bostan et al., 2010, 2013; Buckner, 2013; Buckner et al., 2011; Hoshi et al., 2005; Kelly & Strick, 2003; King et al., 2023; Middleton & Strick, 1994; Strick et al., 2009; Wagner & Luo, 2020). Still, a unifying framework of cerebellar function remains elusive. What specific cognitive computations might the cerebellum implement?

Some recent work on this question focuses on cerebellar contributions to reinforcement learning (RL; Hull, 2020; Kostadinov & Häusser, 2022). Studies in rodents and nonhuman primates have revealed cerebellar activity encoding core components of RL, including reward processing, reward expectation, and reward prediction errors (Heffley et al., 2018; Kostadinov et al., 2019; Sendhilnathan et al., 2020; Wagner et al., 2017). Such signals have been observed in granule cells (Wagner et al., 2017), climbing fiber inputs (Heffley & Hull, 2019; Kostadinov et al., 2019; Larry et al., 2019), and Purkinje cells (Sendhilnathan et al., 2020), especially in areas spanning

lobules Crus I and Crus II (Heffley & Hull, 2019; Sendhilnathan et al., 2020). Moreover, lesion and stimulation studies point to a causal role for the cerebellum in RL (Carta et al., 2019; Huvermann et al., 2025; McDougle et al., 2016; Nicholas et al., 2023). There is now general consensus that the cerebellum supports RL (Manto et al., 2024) in conjunction with circuits in the cerebrum.

Still, key gaps remain, especially in human neuroscience. First, there are practical issues: full coverage of the cerebellum with human functional neuroimaging is rarely achieved (Wang et al., 2025), and cerebellar BOLD signals are susceptible to motor and physiological confounds (Diedrichsen et al., 2010; Schlerf et al., 2012; Striemer et al., 2015). Second, many RL tasks involve sensorimotor contingencies that blur cognitive and motor components of RL (O'Doherty et al., 2003; Sendhilnathan et al., 2020; Seymour et al., 2004). Finally, specific anatomical localization of RL signals in humans is rarely reported when they are incidentally detected, complicating comparisons to model organisms. These challenges may explain heterogeneity in previously reported cerebellar activations during RL (Garrison et al., 2013; Kruithof et al., 2023; O'Doherty et al., 2003).

Theoretical gaps also remain, particularly regarding constraints on cerebellar RL computations. One candidate constraint is temporal sensitivity: In associative motor learning, the cerebellum is optimized for processing rapid cue-outcome intervals (i.e., < ~2 seconds; (Avraham et al., 2022; Brudner et al., 2016; Chettih et al., 2011; Ivry & Keele, 1989; Kitazawa et al., 1995; Smith, 1968), and learning is attenuated when these intervals are exceeded (Gerwig et al., 2008). In contrast, areas like basal ganglia, hippocampus, and neocortex can associate events over longer timescales (Buhusi &

Meck, 2005; Foerde & Shohamy, 2011). We thus hypothesize that the cerebellum may be selectively engaged in RL when predictive associations unfold rapidly.

We tested this hypothesis using choice behavior, eye-tracking, physiological recordings, and model-based fMRI in humans. Our task avoided low-level sensorimotor associations and we optimized cerebellar coverage. By manipulating delays between choice and feedback, we tested temporal constraints on cerebellar RL signals. Moreover, we used whole-brain fMRI to compare cerebellar RL signals to those in canonical RL regions, to link neural responses to behavior, and to ask if RL feedback drives functional correlations with other regions of the brain.

Methods

Participants

Thirty-five individuals participated in the study (Mean age = 22.9 [18-33]; N female = 24). We planned on excluding participants that were not engaging in the task (not responding to >25% of trials or only responding with one response >90% of trials) or that produced too much head motion during the scan (maximum motion over 2.5mm). No participants met our behavioral exclusion threshold. Two participants were excluded for excessive head motion and an additional subject was excluded for technical issues with the data. Our final sample was 32 participants (Mean age = 22.8 [18-33]; N female = 24). The task protocol was approved by the Yale Institutional Review Board. Participants were compensated $30/hr. They were also told that they could receive up to $10 bonus payment for their performance in the task to connect the reward feedback during the task to actual reward. We gave all participants the full bonus payment.

Experimental session

The experimental session was 2hrs long. Participants arrived and were consented and screened for metal. In the scanner, they completed three runs of a probabilistic reinforcement learning (RL) task and three runs of a statistical learning (SL) task (not discussed here). The RL task was always performed before the SL task. Each functional run began with an eye-tracker calibration phase (Eyelink 1000 Plus, Long-range monocular). After the functional runs, we collected a T1 weighted anatomical image before bringing them out of the scanner. Finally, the session concluded with a test phase for the SL task and a debriefing questionnaire. Task and protocol details are explained below.

Reinforcement learning task

Participants performed three runs of a probabilistic reinforcement learning task in the scanner. During each run, participants saw four pairs of images (8 unique images per run) across trials (**Supplemental Figure 1**). On each trial, they saw one pair of images side-by-side and used an MR-safe, two-button button box with their right hand to select one of the two images (index finger to select left image, middle finger to select right image). They received probabilistic reward feedback (reward: +1 or nonreward: +0) on their response after a delay (Figure 1a). One image in each pair was associated with a .25 reward probability and the other with a .75 reward probability throughout a run of trials. Participants were instructed to select the stimulus on each trial that they thought was most likely to give them points. They were informed that their point total would be translated into a monetary bonus at the end of the session (up to $10) to motivate them to perform well and ensure that the point-feedback was rewarding. Crucially, the

location of each image was counterbalanced across trials so that reward information was associated with the image – the abstract choice – and not with a specific location or motor action.

We manipulated the delay between stimulus offset and the presentation of the reward feedback in order to test the hypothesis that the cerebellum would be particularly involved in processing short-latency feedback signals. Participants viewed the pairs of stimuli for 1.5s on each trial, during which time they made their response. Once they made their response, the fixation cross changed color to indicate that their choice had been registered and the stimuli remained on the screen for the remainder of the 1.5s. The stimuli then disappeared and there was a delay before they were presented with reward feedback for 1s. To test the hypothesis that the cerebellum would be sensitive to short-latency feedback, we assigned two pairs in each run to a short feedback delay where the feedback appeared .8s after the stimulus disappeared and the other two pairs to a long feedback delay where feedback appeared after 3s. If they did not respond quickly enough, they received the feedback “Please respond faster.” There was a jittered inter-trial interval (ITI; 1-3s) after feedback presentation.

Prior to the three runs of the task, participants executed a short practice block with two pairs of stimuli to orient them to the task. There was no variation in the delay between responses and feedback in the practice block. During the main runs, participants performed 96 trials during a run (24 presentations per pair randomly interleaved). New stimuli were used in each run to prevent participants from using learned values from previous runs.

Behavioral and physiological analysis

We used participant choices and RTs to characterize their behavior. Trials where participants did not respond were excluded from analysis. Learning was operationalized as the probability of selecting the stimulus associated with the higher reward probability across trials. To quantify learning we implemented a generalized linear model that used iteration (how many times the participant had viewed a specific pair of stimuli) to predict whether or not they chose the higher value stimulus (1 = chose higher value stimulus; 0 = chose lower value stimulus). Overall learning was measured as the probability of selecting the higher value stimulus across all trials. We used t-tests to compare overall learning and RTs between short and long delay trials and to compare gaze and biophysiological measures on rewarded versus nonrewarded trials.

We additionally collected eyetracking data during the entirety of the scan using a long-range Eyelink 1000 plus. We did not obtain eyetracking data from two participants who used MR-safe glasses during the scan which interfered with tracking and an additional participant due to equipment malfunction. We began each run with a calibration phase. We were primarily interested in differences in gaze behavior on rewarded versus nonrewarded trials that might impact our results. To examine this possibility, we summarized gaze behavior within each trial and at feedback by counting the number of saccades in these time intervals and used a t-test to compare between rewarded and nonrewarded trials.

Finally, we collected pulse and respiration data from each participant using Siemens peripheral devices during scanning (samples every 2.5ms). We used a pulse-oximeter positioned on the index finger of the participant's left hand to collect pulse data

and a respiration belt positioned around the participant's abdomen to track respiration. A portion of the data was lost due to signal drop out during the task, or subsequent data storage issues. For five additional participants, we excluded their pulse data from our analyses as the pulse-oximeter was unable to acquire a reliable pulse signal. Thus, of the usable runs of functional BOLD data (86 runs across 32 participants), we retained 74.4% of the respiration data (64 runs across 24 participants) and 55.8% of the pulse data (48 runs across 18 participants) for analysis.

We were primarily interested in whether these physiological metrics varied significantly across rewarded and nonrewarded trials in a way that might impact our results. To test this, we calculated trialwise pulse and respiration rates and then compared these rates on rewarded versus nonrewarded trials. After extracting the data for each metric in each run, we used the *peakdet* function in Matlab (Eli Billauer, 2009) to locate local minima and maxima in the traces. We then calculated a trialwise rate by counting the number of maxima and minima from the onset of the stimuli on one trial to the onset of the stimuli for the next trial and divided that count by the duration between these two time points (in minutes) to determine the rate. We opted to average the number maxima and minima during the time-window-of-interest in this calculation in order to account for situations where a portion, but not all, of a cardiac or respiratory cycle fell within the window of interest. We then calculated subjectwise averages of these rates for rewarded and nonrewarded trials separately and used a t-test to compare the rates.

Computational modeling

We conducted an RL modeling analysis on subjects' trial-by-trial behavioral data. These models allowed us to bridge our behavioral and neural data through the lens of reinforcement learning theory (Sutton & Barto, 1998). Specifically, we aimed to model subject behavior to estimate latent variables that reflect the core prediction error processes of RL, and then use these computational estimates to explain variance in neural data using model-based fMRI (Cohen et al., 2017). This computational analysis thus allowed us to generate inferred reward prediction error (RPE) time-courses for use in our neural analyses.

To model RL behavior and RPEs, we employed a well-established RL model with the basic form:

(1) $Q(s)_{t+1} = Q(s)_t + \alpha\delta$

(2) $\delta = r - Q(s)_t$

where the value (Q) of a given stimulus (s) that is chosen by the participant on trial t is updated according to the RPE (δ) on that trial (the difference between the expected value Q, and received reward r), with a learning rate parameter α.

We compared the fit of three models to participant data. We fit one variant of the model with a single learning rate:

(3) $Q(s)_{t+1} = Q(s)_t + \alpha\delta$

We fit an additional model with separate learning rates (α^+, α^-) for rewarded versus nonrewarded outcomes (Frank et al., 2007; Trach et al., 2025):

(4) $$Q(s)_{t+1} = \begin{cases} Q(s)_t + \alpha^+\delta & if\ r = 1 \\ Q(s)_t + \alpha^-\delta & if\ r = 0 \end{cases}$$

Finally, because our task consisted of stimulus *pairs*, it is possible that participants may make counterfactual inferences, such that if one stimulus of the pair was rewarding them that the other would be unlikely to give them reward, and vice versa (Boorman et al., 2011). In other words, they may update both the Q value of the chosen stimulus, *s*, and of the unchosen stimulus, *s'*, when they receive feedback about *s*. We operationalized this by adding two additional free parameters that dictated how much feedback about the chosen stimulus would impact the Q value of the unchosen stimulus on rewarded and non-rewarded trials:

(5) $$Q(s)_{t+1} = \begin{cases} Q(s)_t + \alpha^+ \delta & if\ r = 1 \\ Q(s)_t + \alpha^- \delta & if\ r = 0 \end{cases}$$

(6) $$Q(s')_{t+1} = \begin{cases} Q(s')_t + \omega^+ (0 - Q(s')_t) & if\ r = 1 \\ Q(s')_t + \omega^- (1 - Q(s')_t) & if\ r = 0 \end{cases}$$

Across all models, action selection between the two presented stimuli was modeled using the softmax function:

(7) $$p(s) = exp(\beta Q(s)) \,/\, \Sigma_i\, exp(\beta Q(s_i))$$

where β reflects the inverse softmax temperature. We used the MATLAB function fmincon to fit our model to each subject's observed choice data, optimizing the parameter values to maximize the log posterior probability of the choice data given the model. During fitting, α, (α^+, α^-), and (ω^+, ω^-) were constrained on [0,1] and β on [0,50], and a Gamma (2,3) prior distribution was used to discourage extreme values of β (Leong et al., 2017; Trach et al., 2025). Importantly, we fit one of each fitted parameter per subject rather than fitting new values for each run (i.e., three of each per participant). The fitted parameters are thus optimized to best reflect participant learning

and performance across runs. This approach gives us more data to derive a stable estimate of these parameters for each subject. Further, theoretically, these fitted parameters should reflect psychological variables that remain relatively stable within the task context, and fitting runs separately could lead to overfitting. The fitting procedure was conducted 200 times per subject with randomized starting parameter values to avoid local minima. Simulated choice data (from the optimized model) were produced to investigate and visualize the model's ability to replicate the main behavioral findings, and to generate the trial-by-trial RPE time courses that were fit to the neural data (see below; see also **Supplemental Figure 2**).

We compared models using the Bayesian Information Criterion (BIC). This fit metric was implemented here as it balances both goodness of fit with model complexity. Smaller values of BIC are indicative of better model fit. We found that the dual learning rate model was the best fit for participant behavior (one learning rate model: summed BIC = 10,291; dual learning rate model: summed BIC = 9,953; inference model: summed BIC = 10,027; **Supplemental Figure 2**). Thus, we used the dual learning rate model to simulate RPEs for our analyses, though we note that all models produced similar RPE timecourses (**Supplemental Figure 2**).Thus, we used the dual learning rate model in our analyses. Simulated choice data (from the optimized model) were produced to investigate and visualize the model's ability to replicate the main behavioral findings, and to generate the trial-by-trial RPE time courses that were fit to the neural data (see below).

fMRI acquisition and pre-processing

MR data were acquired with a 3T Siemens Prisma scanner with a 64-channel head coil at BrainWorks at the WuTsai Institute at Yale University. Whole-brain functional images were collected using an echo-planar imaging sequences with Repetition time (TR) = 1000ms, Echo time (TE) = 30ms, voxel size = 2.5x2.5x3 mm^3, field-of-view = 20.8 x 20.8 x 20.8 cm^3, 48 slices, P to A phase encoding direction, with multi-band acceleration factor = 3 (interleaved) and in-plane acceleration factor = 2. Gradient echo field maps were acquired to correct for distortions due to B0 inhomogeneities (acquisition parameters: voxel size = 3 × 3 × 3 mm^3, field-of-view = 24 × 24 × 24 cm^3). We also collected a high-resolution T1 weighted MPRAGE, voxel size = 1 mm^3, field-of-view = 25.6 × 25.6 × 25.6 cm^3. As previously noted, we used Siemens peripherals to track pulse and respiration during the functional runs as well.

Anatomical data preprocessing

Preprocessing was performed using fMRIPrep 23.2.1 (Esteban et al., 2018, 2019), which is based on Nipype 1.8.6 (Gorgolewski et al., 2011; Krzysztof J. Gorgolewski et al., 2018). The T1w image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.5.0 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL, RRID:SCR_002823 (Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2, RRID:SCR_001847, (Dale et al., 1999), and the brain mask

estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.5.0), using brain-extracted versions of both T1w reference and the T1w template.

Functional data preprocessing

For each of the BOLD runs per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of fMRIPrep, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL, Jenkinson et al., 2002). The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with twelve degrees of freedom to account for distortions remaining in the BOLD reference. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by (Power et al., 2014). The three global signals are extracted within the

CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. (2007) in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5

standardized DVARS were annotated as motion outliers. Additional nuisance time series are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by Patriat et al. (2017). All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using nitransforms, configured with cubic B-spline interpolation. (Copyright Waiver: Much of the above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts unchanged. It is released under the CC0 license.)

Additional pre-processing after fMRIprep

The primary aim of this study was to isolate neural correlates of RL within the cerebellum. Thus, we ran planned analyses on masked cerebellar data, in addition to the whole brain. To constrain analyses within the cerebellum and avoid bleed-over from adjacent visual/temporal regions, we masked the cerebellum from functional data aligned to standard MNI space prior to spatial smoothing. Whole-brain and cerebellar-masked data was spatially smoothed with a 5mm kernel. We excluded runs where participants produced any movement over 2.5mm (1 voxel; N = 10 runs) and excluded participants if they did not have multiple runs to contribute (N = 2). One additional participant was excluded due to technical difficulties.

fMRI analysis

Whole-brain and cerebellar univariate analysis

We modeled BOLD responses using generalized linear models (GLMs) implemented in FSL (version 6.0.7.9). Results from three GLMs are presented in the main text, with additional results from another GLM in the supplement. First, we used a GLM to examine responses to reward versus nonreward feedback. We modeled events as boxcar regressors that lasted for the duration of the event of interest. Regressors of interest were convolved with a double gamma HRF using FSL. Here, we included separate task regressors, in addition to our confound regressors, for 1) choice stimulus appearance, 2) participant response (or delay onset if there was no response), 3) short-delay feedback (with rewarded trials modeled as 1 and nonrewarded trials as -1), and 4) long-delay feedback (with rewarded trials modeled as 1 and nonrewarded as -1). This design meant that main effects of the feedback regressors would highlight regions of the brain that preferentially respond to rewarding versus nonrewarding feedback. Thus, combinations and contrasts of these two regressors can be used to highlight regions of the brain that are sensitive to reward across delays and also regions that are delay-sensitive.

We used a model-based approach to locate neural correlates of RPEs and surprise (|RPE|) in the brain. To do this, we fit RL models to participant choice data (described above) to obtain participant-specific learning and choice parameters. Then, we used these fitted parameters to simulate trial-by-trial RPEs based on the series of trials that each participant experienced. We included separate parametric regressors for RPEs on short-delay trials versus long-delay trials. In the main version of this model (presented in the main text), we included only rewarded trials in the analysis. Crucially,

this allows us to measure true RPE responses that are not confounded with valence. In this model, we additionally included regressors for 1) stimulus appearance, 2) participant response, and 3) the appearance of the feedback. Thus, positive correlations with the RPE regressors should not be due to the appearance of feedback alone or generalized valence responses, but rather with the trial-by-trial variation in inferred RPE. In another version of this model (presented in **Supplemental Figure 3**), we included both rewarded and nonrewarded trials. Here, we included an additional regressor for valence to ensure that correlations with the RPE regressors are not driven by general reward versus nonreward effects (though we note that the valence and RPE regressors are highly colinear, making this model suboptimal). We took a similar approach to examine correlates of surprise. In this case, the parametric regressor was the absolute value of RPEs. Finally, we also leveraged this computational framework to conduct an exploratory analysis examining reward anticipation signals in cerebellum. In this analysis, we used a parametric regressor that had the Q-value of the chosen stimulus for each trial, time-locked to when the shapes were initially presented during the trial. All parametric regressors were z-scored before running the models.

We also conducted a whole-brain GLM with task performance as a covariate in the model to assess if RPE signals were related to behavior. Our primary goal was to assess whether cerebellar RPE activity was related to learning for the short-delay pairs. To that end, we quantified performance for the short-delay pairs by calculating the probability of selecting the higher reward probability shape across all presentations of these pairs. We included this metric as a covariate in the group-level analysis for temporally sensitive RPE signals (i.e., the contrast of RPE activity on rewarded trials on

short versus long delay trials). We repeated this analysis with performance for long delay pairs and again for the difference in performance between short and long delay pairs.

In addition to our regressors of interest, we included a variety of motion and noise regressors in all GLMs. We extracted motion regressors from the fMRIprep pipeline and included six rigid body regressors, a DVARS regressor to account for overall motion, and regressors to scrub high motion (>.5mm) TRs. We included the first 10 components of the aCompCor regressors from fMRIprep to account for noise from white matter or cerebrospinal fluid.

Runs were combined within-subject before group-level analyses (FLAME 1). We used a cluster-forming threshold of $p < .001$ and a family-wise error cluster-corrected threshold of $p < 0.001$, unless otherwise specified in the text. We additionally present results thresholded at 0 in **Supplemental Figure 4.** For cerebellar-specific analyses we cluster-corrected within the cerebellum. We used nilearn (version 0.8.1) to visualize whole-brain results and the SUIT toolbox (Diedrichsen, 2006; Diedrichsen et al., 2009; Diedrichsen & Zotow, 2015) to visualize cerebellar results on flat maps.

ROI analyses

We conducted additional analyses in *a priori* anatomical ROIs to examine correlates of reward processing and RPE. We used anatomical ROIs from the Harvard-Oxford atlas to create ROI masks. We then extracted subject-wise average $\boldsymbol{\beta}$-parameters from the second-level GLM analysis within each mask for each contrast of interest. We used nonparametric bootstrap hypothesis tests to test whether the

measured effect was significantly different from zero in each ROI. We opted for these nonparametric tests as they are robust to statistical assumptions necessary for standard parametric tests. For each bootstrap iteration, we randomly sampled from the measured values with replacement to create a distribution. We repeated this process 1,000 times and computed p-values and 95% confidence intervals by comparing the distribution means to the null value (zero in this case).

Psychophysiological interaction (PPI) analysis

We used psychophysiological interaction analyses (Gitelman et al., 2003) to examine task-dependent changes in functional correlations between the cerebellum and cerebrum during feedback processing. This approach examines whether functional coupling between a seed region and other regions of the brain is modulated by experimental factors. We first extracted a time series of activation from our two *a priori* cerebellar lobular ROIs (Crus I and Crus II) as the physiological regressors in this analysis. Then, we constructed the psychological regressor that identified feedback epochs in the task versus other trial epochs (e.g., stimulus appearance, delay etc.) which was convolved with a canonical hemodynamic response function. Finally, we created a regressor that reflected the interaction of the physiological and psychological regressors. This regressor reveals regions of the brain that significantly covary with the seed region specifically during the epochs of interest.

Data and code availability statement

Processed data and code will be available upon publication of the manuscript.

Results

Human Crus I and Crus II preferentially respond to reward feedback

Participants (N=32) performed a probabilistic reinforcement learning (RL) task while undergoing fMRI. Participants observed two stimuli on each trial and used a button box to select the stimulus (left or right) that they thought was most likely to yield a reward (**Figure 1a**). Their goal was to win points to earn a monetary bonus at the end of the task. After they made their choice, there was a delay period (short delay: 0.8s; long delay: 3s) before they received probabilistic reward feedback (nonrewarded trials: +0; rewarded trials: +1) which they could use to guide future choices. Participants saw four distinct pairs of stimuli within each run (see *Methods* and **Supplemental Figure 1**). In each pair, one stimulus was associated with a .75 probability of reward and the other with a .25 probability of reward. Importantly, the location of the stimuli was randomized across trials, such that rewards were associated with the abstract stimulus choice rather than a specific motor action or spatial location. Our *a priori* regions of interest (ROIs) were Crus I and Crus II in the cerebellum (following work in nonhuman primates (Sendhilnathan et al., 2020) and rodents (Heffley & Hull, 2019)), and canonical subcortical RL regions in nucleus accumbens (NAc), caudate nucleus (Cd), and hippocampus (Hpc; **Figure 1b**).

Participants performed the RL task well, showing significant learning across trial blocks (GLM: choose correct ~ iteration + (1|subject); iteration: b = 0.026, SE = 0.0037, z = 6.93, p < .001; **Figure 1c;** see also **Supplemental Figure 2**). These learning data were fit with an RL model which was used for later model-based fMRI analyses (see

Methods). Importantly, the plateau in the performance curve does not mean that participants were only actively learning early in the block - their performance plateaued at choosing the more rewarding stimulus on 73.8% of trials in the last five iterations of each stimulus pair, rather than 100% which would be the optimal behavior. Additionally, because reward feedback was probabilistic, participants will continue to experience prediction errors throughout the block.

At the group level, feedback delay did not significantly affect choice behavior or reaction times (p(choose higher value stimulus): $t(31) = 0.51$, 95% CI = [-0.02, 0.04], $p = .616$; RT: $t(31) = 0.03$, 95% CI = [-0.01, 0.01], $p = .976$; **Figure 1d**). Crucially, there was comparable gaze behavior (**Figure 1e**; at feedback: N saccades: $t(28) = 0.76$, 95% CI = [-0.05, 0.11], $p = .451$) and physiological measures (**Figure 1f;** rewarded versus nonrewarded trials: respiration: $t(23) = -0.95$, 95% CI = [-0.37, 0.14], $p = .353$; pulse: $t(18) = 1.27$, 95% CI = [-0.25, 0.83], $p = .274$) on rewarded versus nonrewarded trials, meaning that our key neural analyses would not be significantly confounded by these factors. (This is an important control, as confounding factors like eye movements can influence cerebellum activity.)

For our neural analyses, we first used generalized linear models (GLMs) to examine activity between rewarded and nonrewarded trials. In addition to canonical reward processing regions (**Figure 2a, left panel**), we also observed significant activity in the lateral cerebellum around the Crus I/Crus II boundary that reflected greater feedback responses on rewarded versus nonrewarded trials (**Figure 2a, right panel**). The localization of these signals broadly corresponds with previously reported effects in

513 both rodents (Heffley & Hull, 2019) and nonhuman primates (Sendhilnathan et al.,
514 2020).

515

516 *Cerebellar reward signals are temporally constrained*

517 We next tested whether these cerebellar signals were affected by feedback
518 delay. We hypothesized that the cerebellum would be specifically responsive to
519 feedback on short-delay (0.8s) versus long-delay (3s) trials, echoing temporal
520 constraints on cerebellar processing in motor learning tasks (Brudner et al., 2016;
521 Cheng et al., 2008; Kitazawa et al., 1995). We thus examined responses to rewards at
522 short- and long-delay intervals separately. In line with our hypothesis, we found
523 significant cerebellar responses to rewards on short-delay, but not long-delay, trials
524 (**Figure 2b;** see **Supplemental Figure 5**). Activity was primarily localized to Crus I and
525 Crus II. The null findings on long delay trials held even with a relaxed statistical
526 threshold (cluster-corrected in cerebellum, $p < .05$), indicating that the lack of significant
527 reward responses at long delays was not simply a result of conservative thresholding.

528 We additionally asked whether this temporal sensitivity was specific to the
529 cerebellum, or was also reflected in other reward processing regions (NAc, Cd, and
530 Hpc; (Foerde & Shohamy, 2011)). We extracted β-values for the rewarded >
531 nonrewarded contrast (**Figure 2c;** see also **Supplemental Figure 5** for main effects of
532 reward versus non-reward feedback) from these ROIs at short and long delays.
533 Corroborating the whole cerebellum results, the *a priori* anatomical cerebellar ROIs only
534 exhibited significant positive responses at short delays (nonparametric bootstrap test:

short delays: Crus I: M = 0.32, 95% CI = [0.19, 0.45], p < .001; Crus II: M = 0.29, 95% CI = [0.16, 0.42], p < .001; long delays: Crus I: M = -0.09, 95% CI = [-0.26, 0.10], p = .346; Crus II: M = 0.02, 95% CI = [-0.15, 0.19], p = .844). This pattern of results in cerebellum was driven by significant positive activity during feedback on short delay trials (**Supplemental Figure 5;** nonparametric bootstrap test: Crus I: short-delay reward effect: M = 0.51, 95% CI = [0.33, 0.68], p < .001; short-delay no-reward effect: M = 0.28, 95% CI = [0.12, 0.44], p < .001; Crus II: short-delay reward effect: M = 0.31, 95% CI = [0.08, 0.53], p = .006; short-delay no-reward effect: M = 0.07, 95% CI = [-0.10, 0.24], p = .045) and negative activity on long delay trials whose magnitude was invariant to the valence of the feedback (nonparametric bootstrap test: Crus I: long-delay reward effect: M = -0.44, 95% CI = [-0.64,-0.24], p < .001; long-delay no-reward effect: M = -0.43, 95% CI = [-0.59,-0.25], p < .001; Crus II: long-delay reward effect: M = -0.11, 95% CI = [-0.25,-0.05], p = .176; long-delay no-reward effect: M = -0.16, 95% CI = [-0.31,-0.03], p = .010).

In striatum, the caudate showed a similar effect as the cerebellum (nonparametric bootstrap test: short delays: M = 0.20, 95% CI = [0.03, 0.36], p =.020; long delays: M = 0.04, 95% CI = [-0.11, 0.20], p = .648). However, the ventral striatum ROI (NAc) and hippocampus (Hpc) had significant responses at both short and long feedback delays (nonparametric bootstrap test: short delays: NAc: M = 0.84, 95% CI = [0.60, 1.08], p < .001; Hpc: M = 0.23, 95% CI = [0.14, 0.33], p < .001; long delays: NAc: M = 0.64, 95% CI = [0.45, 0.80], p < .001; Hpc: M = 0.25, 95% CI = [0.13, 0.36], p < .001). Only the cerebellar ROIs exhibited *significantly* stronger responses at short versus long feedback delays (short > long: Crus I: M = 0.30, 95% CI = [0.18, 0.44], p <

.001; Crus II: *M* = 0.20, 95% CI = [0.07, 0.33], *p* = .002; Cd: *M* = 0.12, 95% CI = [-0.01, 0.25], *p* = .072; NAc: *M* = 0.15, 95% CI = [-0.01, 0.31], *p* = .072; Hpc: *M* = -0.01, 95% CI = [-0.11, 0.08], *p* = .826). Further, the difference in response to short versus long delay feedback was significantly larger in Crus I than in Cd or Hpc (Crus I: *M* = 0.40; NAc: *M* = 0.21; Cd: *M* = 0.16; Hpc: *M* = -0.02; Crus I vs NAc: 95% CI = [-0.07, 0.44], *p* = .142; Crus I vs Cd: 95% CI = [0.07, 0.41], *p* = .006; Crus I vs Hpc: 95% CI = [0.15,0.68], *p* = 0.002) and in Crus II (*M* = 0.27) than in Hpc (Crus II vs NAc: 95% CI = [-0.2,0.31], *p* = .614; Crus II vs Cd: 95% CI = [-0.1, 0.31], *p* = .298; Crus II vs Hpc: 95% CI = [0.05,0.53], *p* = 0.016). Taken together, these results provide evidence for temporally constrained reward processing in the human cerebellum during RL and also show that the delayed feedback did not globally abolish reward signals.

Model-based analyses reveal time-sensitive RPE computations in human cerebellum

We next tested whether the human cerebellum tracked the core teaching signal in RL: reward prediction error (RPE). We leveraged computational modeling to examine RPEs in the cerebellum (**Figure 3a**). To do this, we fit RL models to each participant's behavior to obtain learning and choice parameters for each subject (see *Methods*; **Figure 1c** and **Supplemental Figure 2**). We then used these subject-specific parameters to simulate trial-by-trial RPEs based on a participant's unique sequence of trials and choices. The time course of these computationally inferred RPEs serves as a prediction about activity in regions of the brain that may encode RPE signals. We restricted our main analyses to RPEs on rewarded trials to avoid the strong collinearity with valence signals (see *Methods;* see **Supplemental Figure 3** for RPE results across rewarded and nonrewarded trials).

We found robust RPE responses at the boundary of Crus I and Crus II in the cerebellum, in addition to canonical RPE signals in the ventral striatum (**Figure 3b**). Like reward-related activation, RPE responses were temporally constrained, and only detectable on short-delay trials (**Figure 3c** and **Supplemental Figure 3**). Regions within Crus I and II, as well as caudate, showed significant RPE responses at short, but not long delays (**Figure 3d**; nonparametric bootstrap test: short delays: Crus I: M = 0.65, 95% CI = [0.47, 0.81], p < .001; Crus II: M = 0.45, 95% CI = [0.28, 0.63], p < .001; Cd: M = 0.36, 95% CI = [0.20, 0.52], p < .001; long delays: Crus I: M = -0.30, 95% CI = [-0.46,- 0.16], p < .001; Crus II: M = 0.01, 95% CI = [-0.10, 0.12], p = .886; Cd: M = -0.10, 95% CI = [-0.27, 0.07], p = .268), while nucleus accumbens and hippocampus displayed robust RPE responses at both short and long delays (short delays: NAc: M = 0.84, 95% CI = [0.60, 1.08], p < .001; Hpc: M = 0.13, 95% CI = [0.04, 0.23], p = .002; long delays: NAc: M = 0.54, 95% CI = [0.39, 0.70], p < .001; Hpc: M = 0.22, 95% CI = [0.11, 0.33], p < .001; although stronger at short delays in NAc: short > long: M = 0.44, 95% CI = [0.23, 0.65], p < .001; Foerde & Shohamy, 2011). The difference in RPE encoding for short versus long delay feedback was significantly larger in Crus I than all other ROIs (Crus I: M = 0.95; Crus II: M = 0.44; NAc: M = 0.3; Cd: M = 0.45; Hpc: M = -0.08; Crus I vs NAc: 95% CI = [0.4, 0.92], p < .001; Crus I vs Cd: 95% CI = [0.26, 0.75], p < .001; Crus I vs Hpc: 95% CI = [0.78, 1.28], p < .001) and for Crus II versus Hpc (Crus I vs NAc: 95% CI = [-0.13,0.43], p = .316; Crus I vs Cd: 95% CI = [-0.28, 0.26], p = 0.99; Crus I vs Hpc: 95% CI = [0.29, 0.77], p < .001). Thus, we found that the human cerebellum encodes prediction errors that are used for learning in a nonmotor learning

task, and that these teaching signals may show similar temporal constraints as those observed in sensorimotor learning.

RPEs can be signed (i.e., sensitive to valence) or unsigned (i.e., only sensitive to outcome surprise, but not valence). Thus, we used the same approach as above to examine neural correlates of surprise (|RPE|). This analysis, however, did not reveal any significant clusters of activity in the cerebellum, even at a relaxed statistical threshold, suggesting that human cerebellar RPE activity is dominated by a reward-sensitive signal (see *Discussion*).

In addition, this model-based analysis framework can be used to examine reward anticipation signals in the brain. Reward anticipation signals can be operationalized as the trial-by-trial Q-value associated with the stimulus that the participant chose. Again, in this exploratory analysis, we identify regions of the brain that have activity that is positively correlated with a computationally inferred value, in this case trial-by-trial Q-values. We did not observe any significant positive activity in cerebellum related to stimulus Q-values, although there was substantial activity in the canonical medial frontal regions that did correlate with stimulus Q-values (**Supplemental Figure 3**; Bartra et al., 2013). We return to this point in the Discussion.

Linking cerebellar RPE activity to behavior

An additional question we asked is whether the apparent cerebellar involvement in RL at short (but not long) feedback delays covaries with behavioral performance in the task. We note that we did not see general performance differences between delay conditions at the group level: On average, participants showed comparable learning and

RTs for stimulus pairs associated with short versus long feedback delays (p(choose higher value): $t(31) = 0.51$, 95% CI = [-0.02, 0.04], $p = .616$; RT: $t(31) = 0.03$, 95% CI = [-0.01, 0.01], $p = .976$; **Figure 1d**). However, we did observe individual differences in learning efficacy.

To investigate brain-behavior correlations, we compared RPE signals for short versus long delay trials in a whole-brain GLM, and included RL performance in the short delay condition as a participant-wise covariate in the model. In other words, this analysis tests for voxels in the brain where stronger temporally sensitive RPE responses covary with how well a participant performed in the short delay condition. Strikingly, this whole-brain analysis revealed significant brain-behavior correlations only within the cerebellum: The analysis revealed a single significant cluster of activity on the Crus I/II boundary (albeit at a slightly relaxed statistical threshold for the whole-brain case; **Figure 4**; see also **Supplemental Figure 4**). As a control, we also repeated this analysis using performance on the long delay condition, and the difference between performance on the short versus long delay conditions, as covariates, and observed no significant clusters. To visualize this result more clearly, we used a mask of voxels that showed this pattern in cerebellum and plotted activity in this region for each subject against their performance at the short delay condition (**Figure 4c**). Although this analysis is correlational, it suggests that better reinforcement learning (in the short delay condition) was related to stronger temporally-sensitive cerebellar RPEs, thus linking our cerebellar results to behavioral outcomes.

Increased connectivity between cerebellar ROIs and cerebrum during feedback

The cerebellum has bidirectional connections to much of the cerebral RL circuitry, including the basal ganglia and prefrontal cortex (Bostan et al., 2013; Bostan & Strick, 2018). Do cerebellar responses to RL feedback reflect connectivity with a wider cerebellar-striatal-frontal functional circuit? To address this question, we asked which regions of the cerebrum might be functionally coupled with the cerebellum during RL, specifically during feedback processing. We conducted exploratory psycho-physiological interaction analyses (PPI; Di et al., 2021; **Figure 5a**) seeded in our two *a priori* anatomical cerebellar ROIs (Crus I and Crus II) to measure if and how functional correlations between the cerebellum and the rest of the brain increase during RL feedback above and beyond other task phases.

We observed significant activity in the medial prefrontal cortex reflecting robust feedback-sensitive functional connectivity with cognitive regions of the cerebellum during RL, albeit at a slightly relaxed statistical threshold in the case of Crus I (**Figure 5b & c**). Crus II also showed RL feedback-related functional correlations with lateral PFC regions. Further, at the relaxed statistical threshold, we saw significant connectivity between Crus II and the caudate nucleus (**Figure 5c**), the subcortical region that showed similar temporal sensitivity for RPE responses. These results suggest that nonmotor reward feedback increases functional coupling between the cerebellum and core corticostriatal circuits that support RL.

Discussion

Here, we present evidence of temporally constrained cerebellar involvement in RL in humans. We found that cognitive regions of the human cerebellum (clusters spanning Crus I and II) responded to rewards and encoded reward prediction errors, extending recent work in model organisms (Heffley & Hull, 2019; Kostadinov et al., 2019; Larry et al., 2019; Sendhilnathan et al., 2020; Wagner et al., 2017). These cerebellar RL responses were sensitive to feedback timing, as in motor learning (Gerwig et al., 2008; Ivry & Keele, 1989; Smith, 1968). Further, the degree of time-sensitive RPE encoding covaried with performance on short delay trials, connecting cerebellar RPEs to the efficacy of RL. Finally, functional connectivity analyses revealed communication between the cerebellum and corticostriatal circuitry during feedback (Bostan et al., 2010; Bostan & Strick, 2018; Carta et al., 2019; Hoshi et al., 2005). These results implicate the cerebellum in prediction error-based learning beyond supervised sensorimotor learning, and highlight the importance of examining cerebellar responses in cognitive neuroscience (Wang et al., 2025).

We note that our results are not the first to report prediction error signals in the human cerebellum (Lee et al., 2025; O'Doherty et al., 2003) or to suggest that the basal ganglia and cerebellum work collaboratively (Bostan & Strick, 2010). Previous work, however, did not necessarily isolate the signals we see here: First, in previous work, rewards were typically attached to specific motor actions, making it unclear whether cerebellar responses were driven by RL versus sensorimotor learning (Sendhilnathan et al., 2024; Wagner et al., 2017). Similarly, in other studies, reward or punishment feedback had significant sensorimotor components, including shocking of the skin and

690 the ingestion of juice (Maschke et al., 2002; O'Doherty et al., 2003). Such feedback is
691 likely to drive sensorimotor learning processes in the cerebellum, and complicates
692 interpretations of cerebellar prediction error responses.

693 The fact that cerebellar RL signals were temporally sensitive could also explain
694 why similar signals have not been reported consistently in the literature. fMRI studies
695 are typically designed with long delays between stimuli to accommodate the slow
696 hemodynamic response or supra-second TR measurements, which could blunt
697 cerebellar learning signals. Additionally, much human fMRI work does not report what
698 exact regions of the cerebellum are involved. Functional mapping of the cerebellar
699 cortex has revealed a mosaic of functional regions, mirroring the complexity of the
700 cerebral cortex (King et al., 2019; Nettekoven et al., 2024; Saadon-Grosman et al.,
701 2024). These advances make specific localization of cerebellar signals increasingly
702 important, and de-emphasize the functional importance of lobular boundaries. Notably,
703 precision mapping research has characterized multiple functional subregions that are
704 interconnected with distinct cortical and striatal networks (Habas et al., 2009;
705 Nettekoven et al., 2024; Saadon-Grosman et al., 2024). The regions of the cerebellum
706 where we see robust RL signals correspond to social-linguistic-spatial and demand
707 regions (Guell et al., 2018; King et al., 2019; Nettekoven et al., 2024). These regions
708 also exhibit resting state connectivity with default mode areas, and wider frontoparietal
709 and corticostriatal networks (Boonstra, 2025; Dobromyslin et al., 2012; Guell et al.,
710 2018; Habas et al., 2009).

711 Importantly, we were able to draw from the animal literature to make broad *a*
712 *priori* predictions about localization. We found correspondence between RL-sensitive

713 regions reported in research with model organisms and our human participants, despite
714 key differences in experimental design. While model organisms afford exceptional
715 access to recording input and output signals in the cerebellum, one advantage of testing
716 humans is that they can engage in rapid RL over a matter of minutes, respond to verbal
717 instructions, and tolerate higher task variability across trials (e.g., multiple feedback
718 delays). We think this work highlights the preservation of cerebellar function across
719 model organisms and emphasizes the importance of crosstalk between human and
720 animal researchers. This idea is particularly relevant as cerebellar research becomes
721 focused on higher-order cognitive functions (Buckner, 2013; Manto et al., 2024; Strick et
722 al., 2009), and as cognitive neuroscientists increasingly consider cerebellar nonmotor
723 functions.

724 One highlight of this work is that we observed consistent evidence of temporal
725 constraints on cerebellar RL feedback processing; cerebellar encoding of reward and
726 RPE was essentially abolished when feedback was delayed. The idea that the
727 cerebellum is primarily involved in sub-second coordination, associative learning, and
728 timing is well-established in the motor domain (Barri et al., 2022; Gerwig et al., 2008;
729 Ivry & Keele, 1989; Kitazawa et al., 1995; Ohmae & Medina, 2015; Smith, 1968). This
730 preference makes sense for a brain region originally adapted for short-timescale
731 sensorimotor control and prediction (Cisek, 2022). At a mechanistic level, properties of
732 granule cell firing and short-term plasticity at mossy fiber-granule cell synapses have
733 been proposed to support sub-second timing computations characteristic of the
734 cerebellum (Barri et al., 2022; Kennedy et al., 2014; Narain et al., 2018; Yamazaki &
735 Tanaka, 2009). These mechanisms can induce timing-specific activity patterns that are

736 then used by Purkinje cells (the main output neurons of the cerebellum) to refine actions
737 or cognitive computations. There are likely biological limits that constrain what time
738 intervals can be precisely learned by these mechanisms (Raymond & Medina, 2018),
739 providing a mechanistic account for why the cerebellum is attuned to short-timescale
740 feedback.

741 Our model-based analysis approach also allowed for exploratory analyses
742 examining reward anticipation signals. Recent work in animal models has found reward
743 anticipation signals during RL in analogous regions (Wagner et al., 2017; Larry et al.,
744 2019; Kostadinov et al., 2019; Heffley & Hull, 2019). For example, Larry et al. (2019)
745 observed climbing fiber responses that scaled with expected reward during the *cue*
746 phase of a RL task (as in classic Temporal-Difference learning; Sutton & Barto, 1998),
747 echoing observations in the domain of eyeblink conditioning (Ohmae et al., 2015).
748 Similarly, Wagner et al. (2017) reported granule cell activity ramping toward an
749 expected reward. In contrast, we did not see significant positive cerebellar activity
750 correlating with trialwise Q-values (**Supplemental Figure 3**). Why the discrepancy
751 between our results and recent work in animal models?

752 The most likely explanation is the major differences in methodological
753 approaches between human fMRI study and electrophysiological approaches. BOLD
754 signals in cerebellum primarily track activity in the mossy fiber-granule cell input
755 pathway to the cerebellum (Shahshahani et al., 2024). Thus, climbing fiber signals at
756 choice may be a relatively weak contribution to the BOLD signal versus combined
757 climbing fiber and mossy fiber-granule cell activity during feedback. As for ramping
758 signals (Wagner et al., 2017), these may be hard to separate from feedback-related

responses given the low temporal resolution of fMRI. Finally, while model-based analyses are extremely useful for making predictions about RPE activity, it is important to remember that these RPE timecourses are inferred from a necessarily simplified model of RL and model fit may vary considerably across participants.

Our connectivity analyses implicate communication between cognitive regions of the cerebellum and neocortical and subcortical regions during RL feedback processing. Some regions of medial PFC also emerged in our reward processing and RPE analyses, perhaps suggesting that co-fluctuations between the cerebellum and these regions are driven by the receipt of rewards. Lateral PFC, in contrast, was not a reward-sensitive region in other analyses. Communication between cerebellum and lateral PFC could reflect general feedback processing, perhaps updating predictions based on new feedback. One caveat is that PPI cannot distinguish between correlations in activity due to functional connectivity versus high correlations in activity driven by alternative sources of shared input (Friston et al., 1997; Sanchez-Romero et al., 2019). The existence of anatomical connections between cerebellum and prefrontal regions via the dentate nucleus and thalamus (Kelly & Strick, 2003) are consistent with a functional coupling explanation, however future work with causal manipulations are necessary to resolve this issue.

Although we are primarily measuring cerebellar input, communication likely involves bidirectional connections between the cerebellum and frontal and subcortical RL regions (Boonstra, 2025; Bostan & Strick, 2010, 2018; Hoshi et al., 2005; Saadon-Grosman et al., 2024; Wagner & Luo, 2020). A recent fMRI paper demonstrated how cerebellar-striatal connectivity might support reward-based motor learning (Lee et al.,

2025); in contrast to our work, they found that cerebellar activity primarily tracked surprise (i.e., unsigned prediction error magnitude). Indeed, we did not find significant activation in the cerebellum tracking surprise. This discrepancy is likely due to significant task differences, where cerebellar signals in Lee and colleagues (2025) were associated with successful versus unsuccessful outcomes of a motor task, not a nonmotor RL task as presented here.

Overall, our results provide evidence of temporally-constrained, behaviorally-relevant RL signals in the human cerebellum. RPE and reward-related signals were prominent in cognitive regions of the cerebellum, corresponding with recent work in animal models. Connectivity results demonstrated functional correlations between functional subregions spanning Crus I and II and regions of the medial and lateral frontal cortex and caudate nucleus during RL. This work extends beyond the well-studied role of the cerebellum in prediction-error based supervised motor learning (Albus, 1971; Hull, 2020b; Marr, 1969; Raymond & Medina, 2018; Wolpert et al., 1998), and contributes to growing evidence that the human cerebellum may support prediction-error based learning across multiple domains.

References

Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, *10*(1–2), 25–61. https://doi.org/10.1016/0025-5564(71)90051-4

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis, Special Issue on The Third International Workshop on Biomedical Image Registration – WBIR 2006*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Avraham, G., Taylor, J. A., Breska, A., Ivry, R. B., & McDougle, S. D. (2022). Contextual effects in sensorimotor adaptation adhere to associative learning rules. *eLife*, *11*, e75801. https://doi.org/10.7554/eLife.75801

Barri, A., Wiechert, M. T., Jazayeri, M., & DiGregorio, D. A. (2022). Synaptic basis of a sub-second representation of time in a neural circuit model. *Nature Communications*, *13*(1), 7902. https://doi.org/10.1038/s41467-022-35395-y

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Boonstra, J. T. (2025). The cerebellar connectome. *Behavioural Brain Research*, *482*, 115457. https://doi.org/10.1016/j.bbr.2025.115457

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. *PLOS Biology*, *9*(6), e1001093. https://doi.org/10.1371/journal.pbio.1001093

820 Bostan, A. C., Dum, R. P., & Strick, P. L. (2010). The basal ganglia communicate with
821 the cerebellum. *Proceedings of the National Academy of Sciences*, *107*(18),
822 8452–8456. https://doi.org/10.1073/pnas.1000496107

823 Bostan, A. C., Dum, R. P., & Strick, P. L. (2013). Cerebellar networks with the cerebral
824 cortex and basal ganglia. *Trends in Cognitive Sciences*, *17*(5), 241–254.
825 https://doi.org/10.1016/j.tics.2013.03.003

826 Bostan, A. C., & Strick, P. L. (2010). The Cerebellum and Basal Ganglia are
827 Interconnected. *Neuropsychology Review*, *20*(3), 261–270.
828 https://doi.org/10.1007/s11065-010-9143-9

829 Bostan, A. C., & Strick, P. L. (2018). The basal ganglia and the cerebellum: Nodes in an
830 integrated network. *Nature Reviews Neuroscience*, *19*(6), 338–350.
831 https://doi.org/10.1038/s41583-018-0002-7

832 Brudner, S. N., Kethidi, N., Graeupner, D., Ivry, R. B., & Taylor, J. A. (2016). Delayed
833 feedback during sensorimotor learning selectively disrupts adaptation but not
834 strategy use. *Journal of Neurophysiology*, *115*(3), 1499–1511.
835 https://doi.org/10.1152/jn.00066.2015

836 Buckner, R. L. (2013). The Cerebellum and Cognitive Function: 25 Years of Insight from
837 Anatomy and Neuroimaging. *Neuron*, *80*(3), 807–815.
838 https://doi.org/10.1016/j.neuron.2013.10.044

839 Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. T. (2011). The
840 organization of the human cerebellum estimated by intrinsic functional

connectivity. *Journal of Neurophysiology*, *106*(5), 2322–2345. https://doi.org/10.1152/jn.00339.2011

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*(10), 755–765. https://doi.org/10.1038/nrn1764

Butcher, P. A., Ivry, R. B., Kuo, S.-H., Rydz, D., Krakauer, J. W., & Taylor, J. A. (2017). The cerebellum does more than sensory prediction error-based learning in sensorimotor adaptation tasks. *Journal of Neurophysiology*, *118*(3), 1622–1636. https://doi.org/10.1152/jn.00451.2017

Carta, I., Chen, C. H., Schott, A. L., Dorizan, S., & Khodakhah, K. (2019). Cerebellar modulation of the reward circuitry and social behavior. *Science*, *363*(6424). https://doi.org/10.1126/science.aav0581

Cheng, D. T., Disterhoft, J. F., Power, J. M., Ellis, D. A., & Desmond, J. E. (2008). Neural substrates underlying human delay and trace eyeblink conditioning. *Proceedings of the National Academy of Sciences*, *105*(23), 8108–8113. https://doi.org/10.1073/pnas.0800374105

Chettih, S. N., McDougle, S. D., Ruffolo, L. I., & Medina, J. F. (2011). Adaptive Timing of Motor Output in the Mouse: The Role of Movement Oscillations in Eyelid Conditioning. *Frontiers in Integrative Neuroscience*, *5*. https://doi.org/10.3389/fnint.2011.00072

Cisek, P. (2022). Evolution of behavioural control from chordates to primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1844), 20200522. https://doi.org/10.1098/rstb.2020.0522

Cohen, J. D., Daw, N., Engelhardt, B., Hasson, U., Li, K., Niv, Y., Norman, K. A., Pillow, J., Ramadge, P. J., Turk-Browne, N. B., & Willke, T. L. (2017). Computational approaches to fMRI analysis. *Nature Neuroscience*, *20*(3), 304–313. https://doi.org/10.1038/nn.4499

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Di, X., Zhang, Z., & Biswal, B. B. (2021). Understanding psychophysiological interaction and its relations to beta series correlation. *Brain Imaging and Behavior*, *15*(2), 958–973. https://doi.org/10.1007/s11682-020-00304-8

Diedrichsen, J. (2006). *A spatially unbiased atlas template of the human cerebellum*.

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). *A probabilistic MR atlas of the human cerebellum*.

Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M., & Ivry, R. B. (2019). Universal Transform or Multiple Functionality? Understanding the Contribution of the Human Cerebellum across Task Domains. *Neuron*, *102*(5), 918–928. https://doi.org/10.1016/j.neuron.2019.04.021

Diedrichsen, J., & McDougle, S. D. (2026). How does the cerebellum contribute to cognitive functions? *PLOS Biology*, *24*(3), e3003688. https://doi.org/10.1371/journal.pbio.3003688

Diedrichsen, J., Verstynen, T., Schlerf, J., & Wiestler, T. (2010). Advances in functional imaging of the human cerebellum: *Current Opinion in Neurology*, *23*, 382–387. https://doi.org/10.1097/WCO.0b013e32833be837

Diedrichsen, J., & Zotow, E. (2015). Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. *PLOS ONE*, *10*(7), e0133402. https://doi.org/10.1371/journal.pone.0133402

Dobromyslin, V. I., Salat, D. H., Fortier, C. B., Leritz, E. C., Beckmann, C. F., Milberg, W. P., & McGlinchey, R. E. (2012). Distinct functional networks within the cerebellum and their relation to cortical systems assessed with independent component analysis. *NeuroImage*, *60*(4), 2073–2085. https://doi.org/10.1016/j.neuroimage.2012.01.139

Drepper, J., Timmann, D., Kolb, F. P., & Diener, H. C. (1999). Non-motor associative learning in patients with isolated degenerative cerebellar disease. *Brain*, *122*(1), 87–97. https://doi.org/10.1093/brain/122.1.87

Eli Billauer. (2009). *Peakdet* (Version 3.4.05) [MATLAB].

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust

preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Ross, B., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., & Isik, A. I. (2018). *fMRIPrep 23.2.1.* [Computer software]. https://doi.org/10.5281/zenodo.852659

Foerde, K., & Shohamy, D. (2011). The role of the basal ganglia in learning and memory: Insight from Parkinson's disease. *Neurobiology of Learning and Memory*, *96*(4), 624–636. https://doi.org/10.1016/j.nlm.2011.08.006

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316. https://doi.org/10.1073/pnas.0706111104

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage*, *6*(3), 218–229. https://doi.org/10.1006/nimg.1997.0291

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *37*(7), 1297–1310. https://doi.org/10.1016/j.neubiorev.2013.03.023

Gerwig, M., Eßer, A. C., Guberina, H., Frings, M., Kolb, F. P., Forsting, M., Aurich, V., Beck, A., & Timmann, D. (2008). Trace eyeblink conditioning in patients with cerebellar degeneration: Comparison of short and long trace intervals.

Experimental Brain Research, *187*(1), 85–96. https://doi.org/10.1007/s00221-008-1283-2

Gitelman, D. R., Penny, W. D., Ashburner, J., & Friston, K. J. (2003). Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. *NeuroImage*, *19*(1), 200–207. https://doi.org/10.1016/S1053-8119(03)00058-2

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Guell, X., Schmahmann, J. D., Gabrieli, J. D., & Ghosh, S. S. (2018). Functional gradients of the cerebellum. *eLife*, *7*, e36652. https://doi.org/10.7554/eLife.36652

Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C. F., Menon, V., & Greicius, M. D. (2009). Distinct Cerebellar Contributions to Intrinsic Connectivity Networks. *Journal of Neuroscience*, *29*(26), 8586–8594. https://doi.org/10.1523/JNEUROSCI.1868-09.2009

Heffley, W., & Hull, C. (2019). Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. *eLife*, *8*, 8–10. https://doi.org/10.7554/eLife.46764

945 Heffley, W., Song, E. Y., Xu, Z., Taylor, B. N., Hughes, M. A., McKinney, A., Joshua, M.,
946 & Hull, C. (2018). Coordinated cerebellar climbing fiber activity signals learned
947 sensorimotor predictions. *Nature Neuroscience*, *21*(10), 1431–1441.
948 https://doi.org/10.1038/s41593-018-0228-8

949 Hoshi, E., Tremblay, L., Féger, J., Carras, P. L., & Strick, P. L. (2005). The cerebellum
950 communicates with the basal ganglia. *Nature Neuroscience*, *8*(11), 1491–1493.
951 https://doi.org/10.1038/nn1544

952 Hull, C. (2020a). Prediction signals in the cerebellum: Beyond supervised motor
953 learning. *eLife*, *9*, e54073. https://doi.org/10.7554/eLife.54073

954 Hull, C. (2020b). Prediction signals in the cerebellum: Beyond supervised motor
955 learning. *eLife*, *9*, e54073. https://doi.org/10.7554/eLife.54073

956 Huvermann, D. M., Berlijn, A. M., Thieme, A., Erdlenbruch, F., Groiss, S. J., Deistung,
957 A., Mittelstaedt, M., Wondzinski, E., Sievers, H., Frank, B., Göricke, S. L., Gliem,
958 M., Köhrmann, M., Siebler, M., Schnitzler, A., Bellebaum, C., Minnerop, M.,
959 Timmann, D., & Peterburs, J. (2025). The cerebellum contributes to prediction
960 error coding in reinforcement learning in humans. *Journal of Neuroscience*.
961 https://doi.org/10.1523/JNEUROSCI.1972-24.2025

962 Iadecola, C., Yang, G., Ebner, T. J., & Chen, G. (1997). Local and Propagated Vascular
963 Responses Evoked by Focal Synaptic Activity in Cerebellar Cortex. *Journal of*
964 *Neurophysiology*, *78*(2), 651–659. https://doi.org/10.1152/jn.1997.78.2.651

965 Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature*
966 *Reviews Neuroscience*, *9*(4), 304–313. https://doi.org/10.1038/nrn2332

967 Ivry, R. B., & Keele, S. W. (1989). Timing Functions of The Cerebellum. *Journal of*
968 *Cognitive Neuroscience*, *1*(2), 136–152.
969 https://doi.org/10.1162/jocn.1989.1.2.136

970 Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for
971 the Robust and Accurate Linear Registration and Motion Correction of Brain
972 Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

973 Kelly, R. M., & Strick, P. L. (2003). Cerebellar Loops with Motor Cortex and Prefrontal
974 Cortex of a Nonhuman Primate. *The Journal of Neuroscience*, *23*(23), 8432–
975 8444. https://doi.org/10.1523/JNEUROSCI.23-23-08432.2003

976 Kennedy, A., Wayne, G., Kaifosh, P., Alviña, K., Abbott, L. F., & Sawtell, N. B. (2014). A
977 temporal basis for predicting the sensory consequences of motor commands in
978 an electric fish. *Nature Neuroscience*, *17*(3), 416–422.
979 https://doi.org/10.1038/nn.3650

980 King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B., & Diedrichsen, J.
981 (2019). Functional boundaries in the human cerebellum revealed by a multi-
982 domain task battery. *Nature Neuroscience*, *22*(8), 1371–1378.
983 https://doi.org/10.1038/s41593-019-0436-x

984 King, M., Shahshahani, L., Ivry, R. B., & Diedrichsen, J. (2023). A task-general
985 connectivity model reveals variation in convergence of cortical inputs to
986 functional regions of the cerebellum. *eLife*, *12*, e81511.
987 https://doi.org/10.7554/eLife.81511

Kitazawa, S., Kohno, T., & Uka, T. (1995). Effects of delayed visual information on the rate and amount of prism adaptation in the human. *The Journal of Neuroscience*, *15*(11), 7644–7652. https://doi.org/10.1523/JNEUROSCI.15-11-07644.1995

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kostadinov, D., Beau, M., Pozo, M. B., & Häusser, M. (2019). Predictive and reactive reward signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. *Nature Neuroscience*, *22*(6), 950–962. https://doi.org/10.1038/s41593-019-0381-8

Kostadinov, D., & Häusser, M. (2022). Reward signals in the cerebellum: Origins, targets, and functional implications. *Neuron*, *110*(8), 1290–1303. https://doi.org/10.1016/j.neuron.2022.02.015

Kruithof, E. S., Klaus, J., & Schutter, D. J. L. G. (2023). The human cerebellum in reward anticipation and outcome processing: An activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews*, *149*, 105171. https://doi.org/10.1016/j.neubiorev.2023.105171

Krzysztof J. Gorgolewski, Oscar Esteban, Christopher J. Markiewwicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, & Dorota Jarecka. (2018). *Nipype* [Computer software]. https://doi.org/10.5281/zenodo.596855

Larry, N., Yarkoni, M., Lixenberg, A., & Joshua, M. (2019). Cerebellar climbing fibers encode expected reward size. *eLife*, *8*, 1–16. https://doi.org/10.7554/eLife.46870

Lee, J. L., Casamento-Moran, A., Bastian, A. J., Cullen, K. E., & Chib, V. S. (2025). Striatal and cerebellar interactions during reward-based motor performance. *Proceedings of the National Academy of Sciences*, *122*(32), e2503373122. https://doi.org/10.1073/pnas.2503373122

Leiner, H. C., Leiner, A. L., & Dow, R. S. (1986). Does the cerebellum contribute to mental skills? *Behavioral Neuroscience*, *100*(4), 443–454. https://doi.org/10.1037/0735-7044.100.4.443

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040

Manto, M., Adamaszek, M., Apps, R., Carlson, E., Guarque-Chabrera, J., Heleven, E., Kakei, S., Khodakhah, K., Kuo, S.-H., Lin, C.-Y. R., Joshua, M., Miquel, M., Mitoma, H., Larry, N., Péron, J. A., Pickford, J., Schutter, D. J. L. G., Singh, M. K., Tan, T., … Yamashiro, K. (2024). Consensus Paper: Cerebellum and Reward. *The Cerebellum*, *23*(5), 2169–2192. https://doi.org/10.1007/s12311-024-01702-0

Marr, D. (1969). A theory of cerebellar cortex. *The Journal of Physiology*, *202*(2), 437–470. https://doi.org/10.1113/jphysiol.1969.sp008820

Maschke, M., Schugens, M., Kindsvater, K., Drepper, J., Kolb, F. P., Diener, H.-C., Daum, I., & Timmann, D. (2002). Fear conditioned changes of heart rate in patients with medial cerebellar lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, *72*(1), 116–118.

McDougle, S. D., Boggess, M. J., Crossley, M. J., Parvin, D., Ivry, R. B., & Taylor, J. A. (2016). Credit assignment in movement-dependent reinforcement learning. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(24), 6797–6802. https://doi.org/10.1073/pnas.1523669113

McDougle, S. D., Tsay, J. S., Pitt, B., King, M., Saban, W., Taylor, J. A., & Ivry, R. B. (2022). Continuous manipulation of mental representations is compromised in cerebellar degeneration. *Brain*, awac072. https://doi.org/10.1093/brain/awac072

Middleton, F. A., & Strick, P. L. (1994). Anatomical Evidence for Cerebellar and Basal Ganglia Involvement in Higher Cognitive Function. *Science*, *266*(5184), 458–461. https://doi.org/10.1126/science.7939688

Narain, D., Remington, E. D., Zeeuw, C. I. D., & Jazayeri, M. (2018). A cerebellar mechanism for learning prior distributions of time intervals. *Nature Communications*, *9*(1), 469. https://doi.org/10.1038/s41467-017-02516-x

Nettekoven, C., Zhi, D., Shahshahani, L., Pinho, A. L., Saadon-Grosman, N., Buckner, R. L., & Diedrichsen, J. (2024). A hierarchical atlas of the human cerebellum for functional precision mapping. *Nature Communications*, *15*(1), 8376. https://doi.org/10.1038/s41467-024-52371-w

Nicholas, J., Amlang, C., Lin, C.-Y. R., Montaser-Kouhsari, L., Desai, N., Pan, M.-K., Kuo, S.-H., & Shohamy, D. (2023). The Role of the Cerebellum in Learning to Predict Reward: Evidence from Cerebellar Ataxia. *The Cerebellum*, *23*(4), 1355–1368. https://doi.org/10.1007/s12311-023-01633-2

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337. https://doi.org/10.1016/S0896-6273(03)00169-7

Ohmae, S., & Medina, J. F. (2015). Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. *Nature Neuroscience*, *18*(12), 1798–1803. https://doi.org/10.1038/nn.4167

Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, *144*, 74–82. https://doi.org/10.1016/j.neuroimage.2016.08.051

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Raymond, J. L., & Medina, J. F. (2018). Computational principles of supervised learning in the cerebellum. *Annual Review of Neuroscience*, *41*, 233–253. https://doi.org/10.1146/annurev-neuro-080317-061948

Saadon-Grosman, N., Du, J., Kosakowski, H. L., Angeli, P. A., DiNicola, L. M., Eldaief, M. C., & Buckner, R. L. (2024). Within-individual organization of the human

cognitive cerebellum: Evidence for closely juxtaposed, functionally specialized regions. *Science Advances*, *10*(45), eadq4037. https://doi.org/10.1126/sciadv.adq4037

Sanchez-Romero, R., Ramsey, J. D., Zhang, K., Glymour, M. R. K., Huang, B., & Glymour, C. (2019). Estimating feedforward and feedback effective connections from fMRI time series: Assessments of statistical methods. *Network Neuroscience*, *3*(2), 274–306. https://doi.org/10.1162/netn_a_00061

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Schlerf, J., Ivry, R. B., & Diedrichsen, J. (2012). Encoding of Sensory Prediction Errors in the Human Cerebellum. *Journal of Neuroscience*, *32*(14), 4913–4922. https://doi.org/10.1523/JNEUROSCI.4504-11.2012

Schmahmann, J. D. (2019). The cerebellum and cognition. *Neuroscience Letters*, *688*(April 2018), 62–75. https://doi.org/10.1016/j.neulet.2018.07.005

Schmahmann, J. D., & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain*, *121*(4), 561–579. https://doi.org/10.1093/brain/121.4.561

Sendhilnathan, N., Bostan, A. C., Strick, P. L., & Goldberg, M. E. (2024). A cerebro-cerebellar network for learning visuomotor associations. *Nature Communications*, *15*(1), 2519. https://doi.org/10.1038/s41467-024-46281-0

Sendhilnathan, N., Ipata, A. E., & Goldberg, M. E. (2020). Neural Correlates of Reinforcement Learning in Mid-lateral Cerebellum. *Neuron*, *106*(1), 188-198.e5. https://doi.org/10.1016/j.neuron.2019.12.032

Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

Shahshahani, L., King, M., Nettekoven, C., Ivry, R. B., & Diedrichsen, J. (2024). Selective recruitment of the cerebellum evidenced by task-dependent gating of inputs. *eLife*, *13*, RP96386. https://doi.org/10.7554/eLife.96386

Smith, M. C. (1968). CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*, *66*(3, Pt.1), 679–687. https://doi.org/10.1037/h0026550

Sokolov, A. A., Miall, R. C., & Ivry, R. B. (2017). The Cerebellum: Adaptive Prediction for Movement and Cognition. *Trends in Cognitive Sciences*, *21*(5), 313–332. https://doi.org/10.1016/j.tics.2017.02.005

Stoodley, C. J. (2016). The cerebellum and neurodevelopmental disorders. *The Cerebellum*, *15*(1), 34–37. https://doi.org/10.1007/s12311-015-0715-3

Stoodley, C. J., Valera, E. M., & Schmahmann, J. D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *NeuroImage*, *59*(2), 1560–1570. https://doi.org/10.1016/j.neuroimage.2011.08.065

Strick, P. L., Dum, R. P., & Fiez, J. A. (2009). Cerebellum and Nonmotor Function. *Annual Review of Neuroscience*, *32*(1), 413–434. https://doi.org/10.1146/annurev.neuro.31.060407.125606

Striemer, C. L., Chouinard, P. A., Goodale, M. A., & de Ribaupierre, S. (2015). Overlapping neural circuits for visual attention and eye movements in the human cerebellum. *Neuropsychologia*, *69*, 9–21. https://doi.org/10.1016/j.neuropsychologia.2015.01.024

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*.

Timmann, D., Drepper, J., Frings, M., Maschke, M., Richter, S., Gerwig, M., & Kolb, F. P. (2010). The human cerebellum contributes to motor, emotional and cognitive associative learning. A review. *Cortex*, *46*(7), 845–857. https://doi.org/10.1016/j.cortex.2009.06.009

Trach, J. E., deBettencourt, M. T., Radulescu, A., & McDougle, S. D. (2025). Rewards transiently and automatically enhance sustained attention. *Journal of Experimental Psychology: General*, *154*(4), 1063–1079. https://doi.org/10.1037/xge0001727

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fMRI studies. *NeuroImage*, *86*, 554–572. https://doi.org/10.1016/j.neuroimage.2013.09.033

Van Overwalle, F., Manto, M., Cattaneo, Z., Clausi, S., Ferrari, C., Gabrieli, J. D. E., Guell, X., Heleven, E., Lupo, M., Ma, Q., Michelutti, M., Olivito, G., Pu, M., Rice, L. C., Schmahmann, J. D., Siciliano, L., Sokolov, A. A., Stoodley, C. J., van Dun, K., … Leggio, M. (2020). Consensus Paper: Cerebellum and Social Cognition. *The Cerebellum*, *19*(6), 833–868. https://doi.org/10.1007/s12311-020-01155-1

Wagner, M. J., Kim, T. H., Savall, J., Schnitzer, M. J., & Luo, L. (2017). Cerebellar granule cells encode the expectation of reward. *Nature*, *544*(7648), 96–100. https://doi.org/10.1038/nature21726

Wagner, M. J., & Luo, L. (2020). Neocortex–Cerebellum Circuits for Cognitive Processing. *Trends in Neurosciences*, *43*(1), 42–54. https://doi.org/10.1016/j.tins.2019.11.002

Wang, B., LeBel, A., & D'Mello, A. M. (2025). Ignoring the cerebellum is hindering progress in neuroscience. *Trends in Cognitive Sciences*, *29*(4), 318–330. https://doi.org/10.1016/j.tics.2025.01.004

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347. https://doi.org/10.1016/S1364-6613(98)01221-2

Yamazaki, T., & Tanaka, S. (2009). Computational Models of Timing Mechanisms in the Cerebellar Granular Layer. *The Cerebellum*, *8*(4), 423–432. https://doi.org/10.1007/s12311-009-0115-7

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.

IEEE Transactions on Medical Imaging, *20*(1), 45–57.
https://doi.org/10.1109/42.906424

Figure legends

Figure 1. *a) Task schematic. Example stimuli pictured in* ***Supplemental Figure 1****. b) Anatomical ROI masks. c) Learning curve and model fit. d) Choice and RT performance on short and long delay trials. e) Gaze on nonrewarded (+0) versus rewarded (+1) trials. Larger values depict data from whole trial and smaller values depict data during the feedback phase only. f) Respiration and pulse rates on rewarded versus nonrewarded trials.*

Figure 2. *a) Contrast of activity on rewarded versus nonrewarded trials at the point of feedback in the whole brain (left panel) and in cerebellum (right panel; plotted on cerebellar flat map, see Methods). Whole brain result is cluster-corrected at p < .001. Cerebellar results are cluster-corrected in cerebellum, p < .001. b) Rewarded > nonrewarded contrast at short and long delays in cerebellum (cluster-corrected in cerebellum, p < .001). c) Contrast of activity for rewarded > non-rewarded trials in cerebellar (Crus I and II), striatal (Nucleus Accumbens, NAc; Caudate Nucleus, Cd), and hippocampal (Hpc) ROIs. ** p < .001; * p < .05*

Figure 3. *a) Schematic of model-based fMRI approach. b) Reward Prediction Error (RPE) responses (whole brain cluster-corrected, p < .001). and cerebellum (cluster-corrected in cerebellum, p < .001). c) RPE response on rewarded trials at short and long feedback delays (cluster-corrected in cerebellum, p < .001). d) RPE responses at short and long delays in cerebellar and other subcortical ROIs. **p < .001; *p < .05*

Figure 4. *a) Whole-brain analysis showing voxels where short RPE > long RPE contrast covaries with performance on short delay trials. (cluster-corrected in whole brain, p < .05). b) Same analysis restricted to cerebellum (cluster-corrected in cerebellum, p < .001). c) Visualization of the correlation between activity in the significant cluster in the cerebellum from panel b, and learning performance for short delay stimuli.*

Figure 5. *a) Illustration of PPI analysis. b) PPI analysis results seeded in Crus I (red regions cluster-corrected at p < .01). c) PPI analysis results seeded in Crus II (red cluster-corrected at p < .01, yellow at p < .001).*