The Cerebellum
https://doi.org/10.1007/s12311-023-01633-2

RESEARCH

The Role of the Cerebellum in Learning to Predict Reward: Evidence from Cerebellar Ataxia

Jonathan Nicholas[1,2] · Christian Amlang[3,4] · Chi-Ying R. Lin[5] · Leila Montaser-Kouhsari[6] · Natasha Desai[3,4] · Ming-Kai Pan[7,8,9] · Sheng-Han Kuo[3,4] · Daphna Shohamy[1,2,10]

Accepted: 2 November 2023

Abstract

Recent findings in animals have challenged the traditional view of the cerebellum solely as the site of motor control, suggesting that the cerebellum may also be important for learning to predict reward from trial-and-error feedback. Yet, evidence for the role of the cerebellum in reward learning in humans is lacking. Moreover, open questions remain about which specific aspects of reward learning the cerebellum may contribute to. Here we address this gap through an investigation of multiple forms of reward learning in individuals with cerebellum dysfunction, represented by cerebellar ataxia cases. Nineteen participants with cerebellar ataxia and 57 age- and sex-matched healthy controls completed two separate tasks that required learning about reward contingencies from trial-and-error. To probe the selectivity of reward learning processes, the tasks differed in their underlying structure: while one task measured incremental reward learning ability alone, the other allowed participants to use an alternative learning strategy based on episodic memory alongside incremental reward learning. We found that individuals with cerebellar ataxia were profoundly impaired at reward learning from trial-and-error feedback on both tasks, but retained the ability to learn to predict reward based on episodic memory. These findings provide evidence from humans for a specific and necessary role for the cerebellum in incremental learning of reward associations based on reinforcement. More broadly, the findings suggest that alongside its role in motor learning, the cerebellum likely operates in concert with the basal ganglia to support reinforcement learning from reward.

Keywords Reward · Reinforcement learning · Cerebellum · Ataxia

Introduction

It is well established that the cerebellum is required for refining movement through supervised motor learning [1–4]. The cerebellum receives error signals from climbing fiber input which then alters Purkinje cell plasticity to adapt motor behavior in service of minimizing future error [5–7]. However, recent findings have challenged the notion that the cerebellum is solely responsible for supervised learning of motor behavior and instead suggest that

✉ Sheng-Han Kuo
sk3295@cumc.columbia.edu

✉ Daphna Shohamy
ds2619@columbia.edu

[1] Department of Psychology, Columbia University, New York, NY, USA

[2] Zuckerman Mind Brain Behavior Institute, Columbia University, Quad 3D, 3227 Broadway, New York, NY 10027, USA

[3] Department of Neurology, Columbia University Medical Center, 650 W. 168th St, Rm 305, New York, NY 10032, USA

[4] Initiative for Columbia Ataxia and Tremor, Columbia University Medical Center, New York, NY, USA

[5] Department of Neurology, Baylor College of Medicine, Houston, TX, USA

[6] Department of Neurology, Stanford University School of Medicine, Palo Alto, CA, USA

[7] Department of Medical Research, National Taiwan University Hospital, 100 Taipei, Taiwan

[8] Department and Graduate Institute of Pharmacology, National Taiwan University College of Medicine, 100 Taipei, Taiwan

[9] Cerebellar Research Center, National Taiwan University Hospital, Yun-Lin Branch, Yun-Lin, Taiwan

[10] Kavli Institute for Brain Science, Columbia University, New York, NY, USA

Published online: 08 December 2023

the cerebellum may also be involved in the processing of reward more generally [8–19]. In particular, climbing fiber inputs to the cerebellum encode expected reward [13, 15, 17, 19], and cerebellar Purkinje cells have been found to report reward-based prediction errors [11, 12, 18]. These signals are essential ingredients for reinforcement learning, or learning that allows an organism to determine from trial-and-error feedback which actions should be taken in order to maximize future expected reward. The presence of reward-related processing in the cerebellum suggests that it may play a role in reinforcement learning alongside its capacity for supervised motor learning [20]. This proposal challenges not only our current understanding of cerebellar function, but also our understanding of how the brain learns from reward more broadly [5, 21].

Although research on the cerebellum's function in reward learning is growing, the vast majority of work has been done in animal models [10–19], and evidence in humans remains limited. Human neuroimaging studies have revealed correlational evidence that the cerebellum is involved in tasks unrelated to movement [22]; however, despite some reports of BOLD activity in the cerebellum in response to reward across several early imaging studies [23–25], more direct investigations of the role of cerebellum in reward-related behaviors in humans are lacking. The aim of the present study was to fill this gap by testing whether individuals with damage to the cerebellum, as occurs in cerebellar ataxia (CA), are impaired in their ability to acquire stimulus-reward associations.

Our study builds upon a rich literature focused on learning about reward from trial-and-error feedback. This process has been studied extensively using models of incremental learning, which rely on error-driven rules that summarize experiences with a running average [26–28]. During reward learning of this type, an agent uses the outcome of a recent decision to associate some stimulus with an action. Following successful learning, actions that are more likely to be rewarded are more likely to be repeated. This simple mechanism has been evoked to explain instrumental conditioning behavior and is well-captured by reward prediction error signals in midbrain dopamine neurons that project to the striatum [27, 29]. This error signal is also precisely what has been implicated in recent animal models of cerebellar contributions to reward learning [9], suggesting an additional, albeit unclear, role for the cerebellum in this process. Whether these cerebellar contributions are actually needed for successful incremental reward learning in humans is at present unknown.

To answer this question, we asked individuals with CA to complete a series of tasks that required them to learn associations between stimuli from trial-and-error feedback in order to maximize expected reward. CA is defined as a lack of coordination caused by disorders that affect cerebellar function [30]. A large variety of conditions can cause CA, ranging from immune-mediated disease to genetic and neurodegenerative disorders. Given the presence of cerebellar dysfunction in CA cases, studying individuals with CA is a common method used to investigate the necessary physiological functions of the cerebellum in humans.

Nineteen individuals with CA and 57 age- and sex-matched healthy controls (HC) completed two tasks (Fig. 1). The first, referred to throughout as the *incremental learning* task, allowed us to measure each participants' ability to learn about reward incrementally. This task was motivated by recent work using a similar simplified paradigm to investigate cerebellar-based incremental learning in non-human primates [10, 11]. The second task, referred to throughout as the *multiple learning strategies* task, allowed us to measure whether any impairments were specific to incremental learning alone. In the multiple learning strategies task, learning about reward can be supported by an alternative strategy based on episodic memory for trial-unique past outcomes. Healthy adults readily use of both of these strategies in this task [31, 32]. We hypothesized that cerebellar dysfunction would lead specifically to impaired incremental reward learning relative to healthy controls.

Materials and Methods

Cerebellar Ataxia Participants

Nineteen individuals with cerebellar ataxia were recruited from the Ataxia Clinic, Columbia University Medical Center and completed both tasks (see Table 1 for information about basic CA participant demographics and diagnoses). Due to hardware issues, data from one participant on each task was not saved. The first CA participant also completed a shorter pilot version of the incremental learning task, and several changes were made before running this task on the other 18 CA participants. Thus, the final sample for the incremental learning task was 17 CA participants, and the final sample for the multiple learning strategies task was 18 CA participants. Task order was counterbalanced. A neuropsychological battery comprising the Montreal Cognitive Assessment (MOCA), Beck's Depression Inventory (BDI), MESA digit forward and backward span, trail making test A and B, and the cerebellar cognitive affective syndrome scale (CCAS) was conducted between tasks for each participant. This battery was specifically selected based on the current understanding of the cerebellum's role and association with non-motor symptoms, such as depression [33], executive function [34, 35], and attention [36]. Patients were compensated at a rate of $15/hour for their time, and the in-person session took an average of 2.5 h, including breaks.

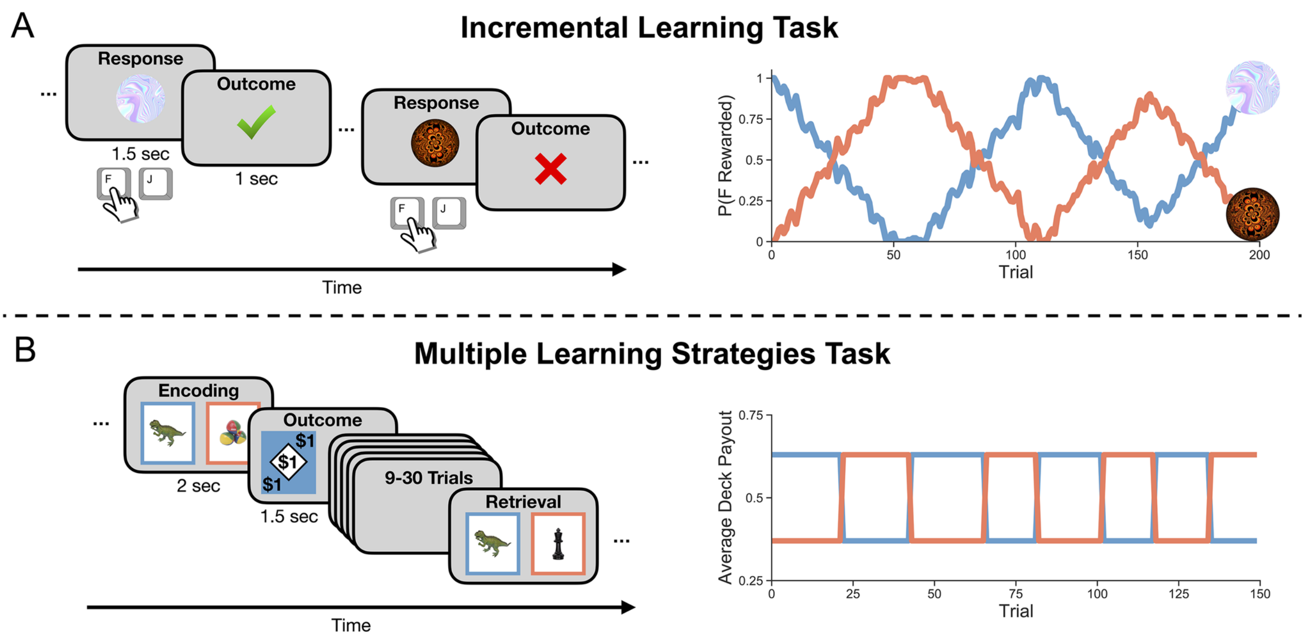

Fig. 1 Design of the incremental learning and multiple learning strategies tasks. **A** Left: Trial design for the incremental learning task. Participants saw one of two fractal cues on the screen and were required to press either the F key with their left hand or the J key with their right hand. Following their choice, they received binary probabilistic feedback about whether they were correct or not. Right: Drifting cue-response-reward contingencies over the course of the incremental learning task. The probability that the F key is rewarded is shown for each cue in blue and orange. **B** Left: Trial design or the multiple learning strategies task. Participants chose between two decks of cards (one blue and one orange) and received an outcome between $0 and $1 in intervals of 20 cents. Each card featured a trial-unique object that could repeat once every 9–30 trials. Participants were told that if they saw the same card again, it would be worth the same amount as the first time that it appeared. Right: An example of how average deck value reversed throughout the course of the multiple learning strategies task

Healthy Controls

Age- and sex-matched participants were recruited through Amazon Mechanical Turk using the Cloud Research Approved Participants feature [37]. To account for potential variability due to online data collection, three matched controls were collected for each CA participant, bringing the total number of controls to 57 (3:1 match). Data from one control was excluded for the multiple learning strategies task due to random responding. Task order was counterbalanced such that the tasks were completed in the identical order to each control's matched CA participant. A modified online neuropsychological battery consisting of 7 measures was completed in between each task for comparison to individuals with CA. Five of these measures (Semantic Fluency, Phonemic Fluency, Category Switching, Similarities, and Go No Go) were directly taken from the CCAS, and two others were composed of the MESA digit forward back backward span. Participant recruitment was restricted to the USA. Before starting each task, all participants were required to score 100% on a quiz that tested their comprehension of the instructions. Controls were compensated at a rate of $15/hour for their time.

Experimental Design

Incremental Learning Task

In the incremental learning task (Fig. 1A), participants were told that they would be playing a game where they were required to press a key, either F or J, whenever one of two symbols was seen, and that they would receive feedback about whether they had pressed correctly following each trial. They were then informed that it was their job to determine which key they should press for each symbol, and that what key is best will change throughout the experiment. Outcomes were determined by a drifting probability such that each button was correct for each image 50% of the time. Critically, these probabilities differed over time, thus encouraging constant learning throughout the task. Participants were told to press the F key with their left index finger and the J key with their right index finger. The response period during which the symbol remained on the screen lasted 1.5 s, with feedback displayed for 1 s immediately following the response period. An intertrial interval featuring a fixation cross was shown for an average of 1 s, but varied between 0.5 and 1.5 s. Lastly, to provide a rewarding outcome for

Table 1 Basic CA participant demographics and neuropsychiatric measures

Participant	Age (years)	Sex	Diagnosis	CCAS	MoCA	BDI	QUIP	FDS	BDS	TMTA (sec)	TMTB (sec)
Participant 1	40	M	SCA3	59	21	15	40	8	5	50	150
Participant 2	33	M	SCA3	77	21	9	12	11	5	21	261
Participant 3	20	F	SCA2	92	27	23	10	13	4	41	71
Participant 4	52	F	MSA-C	85	27	4	0	8	3	23	158
Participant 5	61	F	SCA2	92	23	15	38	8	4	49	169
Participant 6	56	M	MSA-C	100	26	7	8	13	4	66	127
Participant 7	52	M	SCA2	62	26	13	55	11	4	105	287
Participant 8	41	F	SCA2	95	29	18	2	10	8	45	97
Participant 9	43	M	SCA1	87	27	0	6	10	6	52	104
Participant 10	62	F	MSA-C	70	21	4	25	6	6	45	121
Participant 11	54	M	SCA2	72	21	0	5	11	5	32	100
Participant 12	67	F	ILOCA	101	29	12	16	12	5	51	88
Participant 13	60	F	SCA3	104	28	1	0	14	9	42	113
Participant 14	51	F	SCA10	74	25	13	8	8	2	35	83
Participant 15	66	M	SCA1	60	23	4	20	7	3	66	127
Participant 16	49	M	IMCA	86	28	17	6	13	7	81	225
Participant 17	54	M	FA	98	26	24	8	10	8	50	80
Participant 18	33	F	FA	84	27	3	26	9	4	38	84
Participant 19	54	F	IMCA	113	28	18	4	11	8	33	62

SCA spinocerebellar ataxias, *MSA-C* multiple system atrophy, cerebellar type, *ILOCA* idiopathic late onset cerebellar ataxia, *IMCA* immune-mediated cerebellar ataxia, *FA* Friedreich's ataxia, *CCAS* cerebellar cognitive affective/Schmahmann syndrome scale, *MoCA* Montreal cognitive assessment, *BDI* Beck's depression inventory, *QUIP* questionnaire for impulsive-compulsive disorders in Parkinson's disease, *FDS* forward digit span, *BDS* backward digit span, *TMTA* trail making test part A, *TMTB* trail making test part B

correct responses, participants were informed that they could earn bonus money based on their performance. Correct responses were worth an additional cent each, and, on average, healthy control participants earned $1.24, and CA participants earned $0.92 in bonus compensation.

Multiple Learning Strategies Task

The other task completed by participants was previously developed by our lab [31, 32] to measure the relative contribution of incremental learning and episodic memory to decisions (Fig. 1B). Participants were told that they would be playing a card game where their goal was to win as much money as possible. Each trial consisted of a choice between two decks of cards that differed based on their color (red or blue). Participants had two seconds to decide between the decks. The outcome of each decision was then immediately displayed for 1.5 s. Following each decision, participants were shown a fixation cross during the intertrial interval period which varied in length (mean = 1.5 s, min = 1 s, max = 2 s). Decks were equally likely to appear on either side of the screen on each trial. Participants completed a total of 150 trials.

Participants were made aware that there were two ways they could earn bonus money throughout the task, which allowed for the use of incremental learning and episodic memory respectively. First, at any point in the experiment one of the two decks was "lucky," meaning that the expected value (V) of one deck color was higher than the other (V_{lucky}=63¢, $V_{unlucky}$=37¢). Outcomes ranged from $0 to $1 in increments of 20¢. Critically, the mapping from V to deck color reversed periodically throughout the experiment, which incentivized participants to utilize each deck's recent reward history to determine the identity of the currently lucky deck. Second, to assess the use of episodic memory throughout the task, each card within a deck featured an image of a trial-unique object that could re-appear once throughout the experiment after initially being chosen. Participants were told that if they encountered a card a second time it would be worth the same amount as when it was first chosen, regardless of whether its deck color was currently lucky or not. On a given trial t, cards chosen once from trials $t-9$ through $t-30$ had a 60% chance of reappearing following a sampling procedure designed to prevent each deck's expected value from becoming skewed by choice, minimize the correlation between the expected value of previously seen cards and deck expected value, and ensure that choosing a previously selected card remained close to 50¢. Participants were paid a bonus in proportion to their final combined earnings in this task (total earnings/100). On average, healthy control participants earned $0.76, and CA participants earned $0.70 in bonus compensation on this task.

Following completion of the multiple learning strategies task, we tested participants' memory for the trial-unique objects. Participants completed up to 54 three-part memory trials. An object was first displayed on the screen, and participants were asked whether or not they had previously seen the object and were given five response options: Definitely New, Probably New, Don't Know, Probably Old, Definitely Old. If the participant indicated that they had not seen the object before or did not know, they moved on to the next trial. If, however, they indicated that they had seen the object before they were then asked if they had chosen the object or not. Lastly, if they responded that they had chosen the object, they were asked what the value of that object was.

Computational Models

In order to capture subjective estimates of incrementally constructed value on each task, we fit computational models to participants' choices. Below we describe each of these models in detail.

Q Learning Models

We modeled incremental reward learning using a Q Learning model, which is a standard model-free reinforcement learner that assumes a stored value (Q) for each deck is updated over time [26, 28]. Q is then referenced on each decision in order to guide choices. After each outcome, r_t, the value for an option Q_1 is updated according to the following rule if that option is chosen:

$$Q_{1,t+1} = Q_{1,t} + \alpha(r_t - Q_{1,t})$$

And is not updated if a different option is chosen:

$$Q_{1,t+1} = Q_{1,t}$$

Likewise, if a different option is chosen, its value is updated equivalently. Large differences between estimated value and outcomes therefore have a larger impact on updates, but the overall degree of updating is controlled by the learning rate, α, which is a free parameter constrained to lie between 0 and 1.

For the incremental learning task, the model learned separate Q values for each cue and button combination, such that four Q values were estimated in total. Decisions were then modeled using the following rule:

$$P(ChooseF) = \sigma(\beta_{0,1} + \beta_{0,2} + \beta_{1,1}(Q_{F,1} - Q_{J,1}) + \beta_{1,2}(Q_{F,2} - Q_{J,2}))$$

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

such that four inverse temperatures β were estimated to capture a bias toward choosing a key for each cue ($\beta_{0,1}$ and $\beta_{0,2}$) and sensitivity to incrementally learned value for each cue ($\beta_{1,1}$ and $\beta_{1,2}$). This model is referred to as the "Q Learner" model throughout the text.

For the multiple learning strategies task, the model learned separate Q values for each deck color, such that two Q values were estimated in total. Decisions were then modeled using the following rule:

$$P(ChooseRed) = \sigma(\beta_1(Q_R - Q_B) + \beta_2(OldValue) + \beta_3(Old))$$

such that three inverse temperatures β were estimated to capture sensitivity to incrementally learned value (β_1), sensitivity to the value of previously seen objects (β_2), and a bias toward choosing the deck featuring a previously seen object regardless of its value (β_3). The predictor *OldValue* was the coded true value of a previously seen object (ranging from 0.5 if the value was \$1 on the red deck or \$0 on the blue deck to − 0.5 if the value was \$0 on the red deck and \$1 on the blue deck), and the predictor *Old* was coded as 0.5 if the red deck featured a previously seen object and − 0.5 if the blue deck did instead. For both of these predictors, trials that did not feature a previously seen object were coded as 0. This model is referred to as the "Hybrid" model throughout the text.

Biased Responder Model

For both tasks, we compared the performance of the Q Learning models to a model which made choices that were completely independent of reward information. For the incremental learning task, this model was simply:

$$P(ChooseF) = \sigma(\beta_{0,1} + \beta_{0,2})$$

such that choices depended only on choosing a button to press for each cue throughout the experiment. For the multiple learning strategies task, this model was

$$P(ChooseRed) = \sigma(\beta_0)$$

such that choices depended only on preferring one deck over the other throughout the experiment. Our logic in using this model as a baseline was that responses captured by the Q learning models should, at a minimum, outperform a biased responder that did not consider reward in order for it to make meaningful predictions about participants' behavior.

Posterior Inference and Model Comparison

Model parameters for each participant were estimated using Bayesian inference. The joint posterior was approximated using No-U-Turn Sampling [38] as implemented in stan [39]. Four chains with 2000 samples (1000 discarded as burn-in) were run for a total of 4000 posterior samples per model per subject. Chain convergence was determined by ensuring that the Gelman-Rubin statistic $\widehat{R}$ was close to 1 for all parameters. For the incremental learning task, the Q learner did not converge for one CA participant, and so that individual and their matched controls were removed from further model-based analyses. For the multiple learning strategies task, all models for all participants converged.

Under this approach, the likelihood function for all models can be written as

$$c_t \sim Bernoulli(\theta_t)$$

where c_t is 1 if the subject chose F (in the resonse mapping task) or red (in the multiple learning strategies task). Here, θ_t is the linear combination of inverse temperature parameters and predictors explained above for each model. For the Q learning models, the learning rate,α, had the following weakly informative prior:

$$\alpha \sim \beta(0, 1)$$

For all models, every inverse temperature parameter had the following weakly informative prior:

$$\beta \sim \mathcal{N}(0, 5)$$

Model fit was assessed using approximate leave-one-out cross validation estimated using Pareto-smoothed importance sampling [40]. The expected log pointwise predictive density (ELPD) was computed and used as a measure of out-of-sample predictive fit for each model.

Bayesian Observers

In order to provide a normative performance benchmark, we simulated beliefs about incremental value as estimated by Bayesian observers for each task. For the incremental learning task, this learner was a Kalman Filter [41], and for the multiple learning strategies task this learner was a reduced Bayesian change-point detection model [42]. Choices in the incremental learning task were made according to which button the observer believed was the most likely to be rewarded for each cue at each time point. Choices in the multiple learning strategies task were made differently depending on whether a previously seen object was present. For trials in which no previously seen object was shown, the observer responded according to its beliefs about deck value. For trials in which a previously seen object was present, however, the observer compared the value of that object to its belief about deck value for the opposing deck and chose accordingly. In this way, the observer was augmented with "perfect" episodic memory.

Regression Models

Mixed effects Bayesian regressions were used to test effects of group (CA participant or control). Group membership was allowed to vary randomly by CA participant identifier, *pid*, such that CA participants and matched controls were assigned the same ID. In these models, *GroupID* was coded as – 0.5 for CA participants and 0.5 for controls. We additionally controlled for working memory ability by including backwards digit span scores, *dsBwd*, as a standardized covariate in these analyses.

For the incremental learning task, we assessed behavioral performance using the following logistic regression:

$$\begin{aligned} p(Correct) = \sigma(&\beta_0 + b_{0,pid[t]} + GroupID_t(\beta_1 + b_{1,pid[t]}) \\ &+ pFReward1_t\left(\beta_2 + b_{2,pid[t]}\right) + GroupID_t \times pFReward1_t\left(\beta_3 + b_{3,pid[t]}\right) \\ &+ pFReward1_t^2(\beta_4 + b_{4,pid[t]}) + GroupID_t \times pFReward1_t^2(\beta_5 + b_{5,pid[t]}) \\ &+ dsBwd\beta_6 + RT\beta_7) \end{aligned}$$

Here, and in all regressions described in this section, β stands for fixed effects, and b stands for random effects of CA participant ID. The predictor *pFReward*1 indicates the true underlying difficulty of the task and is the probability that the F key was rewarding for cue one. A second-order polynomial was included for this predictor as extreme values indicate portions of the task that are easier and middling values indicate portions of the task that were more difficult. Interaction effects of this predictor and group were included to capture differences in sensitivity to the underlying task difficulty between the groups. Lastly, the reaction time, RT, on each decision was included as a standardized covariate in this analysis to account for any differences that may be due to slowed responding by individuals with CA on this task.

For both the incremental learning and multiple learning strategies tasks, we assessed whether there were differences between the groups on Q learning model performance compared to the baseline biased responder model with the following linear regression:

$$ELPDDifference = \beta_0 + b_{0,pid[t]} + GroupID_t(\beta_1 + b_{1,pid[t]}) + dsBwd\beta_2$$

where *ELPDDifference* was the difference in model performance (Q Learning model ELPD – Biased Responder ELPD; see above) for each subject.

For the multiple learning strategies task, we assessed behavioral incremental learning performance using the following logistic regression:

$$p(ChooseLucky) = \sigma(\beta_0 + b_{0,pid[t]} + T_{-3:3} \times GroupID_t(\beta_{1:7} + b_{1:7,pid[t]}) + dsBwd\beta_8)$$

In this regression, we grouped trials according to their distance from a reversal, up to three trials prior to ($t = -3 : -1$), during ($t = 0$), and after ($t = 1 : 3$) a reversal occurred. We then dummy coded them to measure their effects on the degree to which the lucky deck was chosen and interacted each dummy coded regressor with group to measure how this was affected by group membership.

We also assessed behavioral sensitivity to low- and high-value objects using a similar logistic regression:

$$\begin{aligned} p(ChooseOld) = \sigma(&\beta_0 + b_{0,pid[t]} + LowValue_t \times GroupID_t(\beta_1 + b_{1,pid[t]}) \\ &+ HighValue_t \times GroupID_t\left(\beta_2 + b_{2,pid[t]}\right) + dsBwd\beta_3) \end{aligned}$$

In this regression, in order to assess sensitivity to low- and high-value objects separately, we grouped trials on which a previously seen object was shown according to whether the value of the object was less than 0.5 (*LowValue*) or greater than 0.5 (*HighValue*). We then estimated the effect of the interaction of each of these variables with group membership to measure differences between groups.

We then assessed the degree to which each group used either incrementally learned deck value, the value of previously seen objects, or a bias toward previously seen objects regardless of their value as estimated by the Hybrid Q learning model using a simple linear regression of the following form for each of these inverse temperature parameters and groups:

$$InvTemp_s = \beta_0 + dsBwd\beta_1$$

Here we interested primarily in the intercept, β_0, as this determined the degree to which each group's inverse temperatures were above zero. We additionally assessed differences between groups on each of these measures by including fixed and random effects for group that varied by matched participant ID, as in previously described regression analyses.

We also assessed the impact of group on subsequent memory performance following the multiple learning strategies task using the following linear regression:

$$Dprime = \beta_0 + b_{0,pid[t]} + GroupID_t(\beta_1 + b_{1,pid[t]}) + dsBwd\beta_2$$

where *Dprime* is the signal detection measure $d\prime$, which is the difference in z scored hit rate and false alarm rate for each participant. The hit rate was calculated as the proportion of trials on which a participant answered "Old" when the presented object was previously seen (including both "Definitely Old" and "Probably Old" responses). The false alarm rate was calculated as the proportion of trials on which a participant answered "New" when the presented object was not previously seen (including both "Definitely New" and "Probably New" responses). Trials on which a participant responded "Don't Know" were dropped from further analysis.

We were also interested in determining whether there were any differences in reaction times between individuals with CA and matched controls due to motor impairment. For both tasks, we did this by assessing whether there were any differences in reaction time between groups:

$$RT = \beta_0 + b_{0,pid[t]} + GroupID_t(\beta_1 + b_{1,pid[t]})$$

where *RT* was the median reaction time across trials in either task. A separate regression of this form was used for each of the two tasks. We also assessed whether there were differences on each neuropsychological measure using a similar regression.

For all regression analyses, fixed effects are reported in the text as the mean of each parameter's marginal posterior distribution alongside 95% credible intervals, which indicate where 95% of the posterior density falls. Parameter values outside of this range are unlikely given the model, data, and priors. Thus, if the range of likely values does not include zero, we conclude that a meaningful effect was observed.

Results

Impaired Reward Learning in the Incremental Learning Task

Our first goal was to assess CA participants' baseline ability to learn incrementally from reward using the incremental learning task. On this task, CA participants made overall fewer correct choices compared to healthy controls

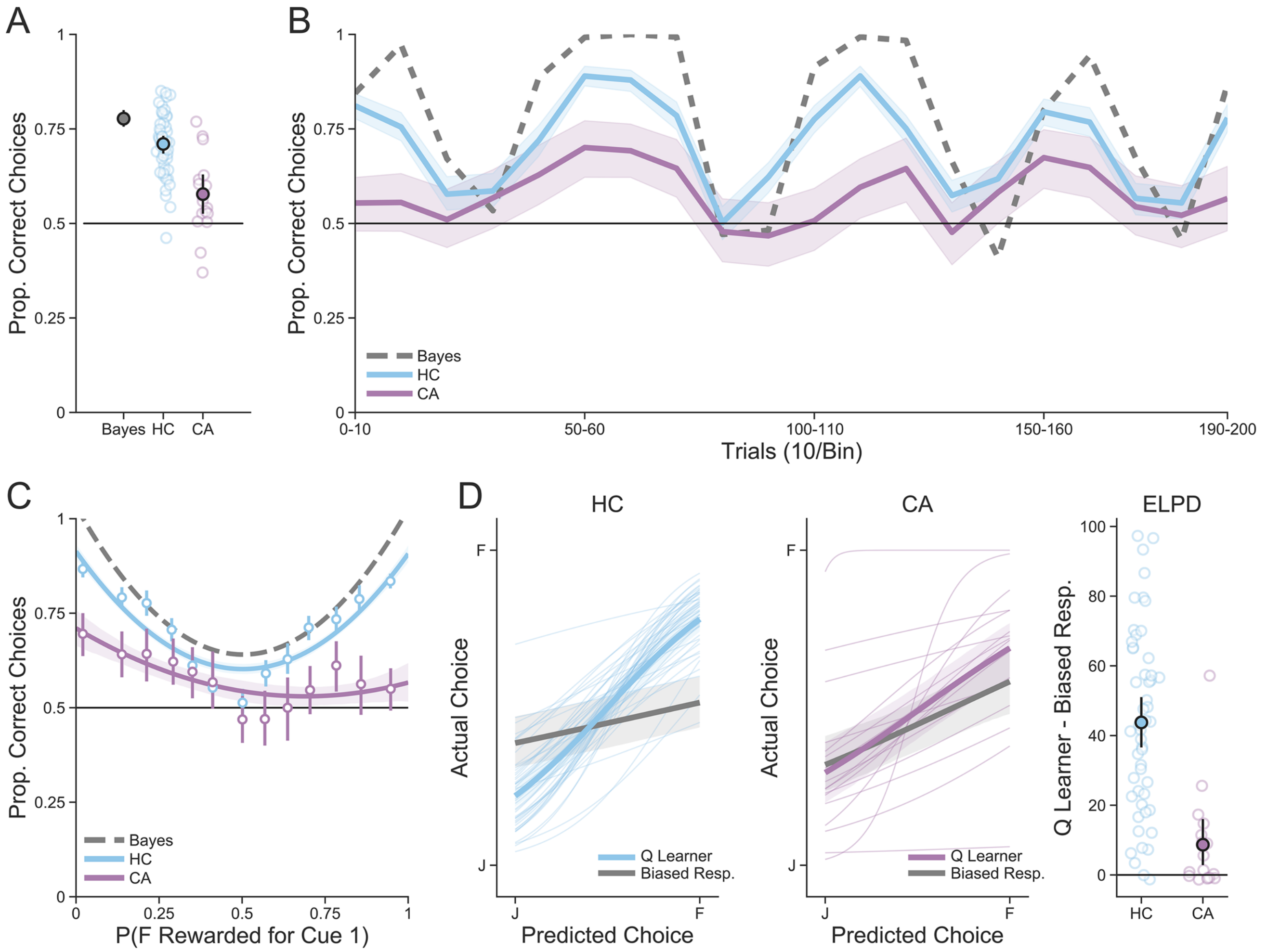

Fig. 2 Performance on the incremental learning task. **A** Performance on the incremental learning task averaged across all trials for healthy controls (HC) and CA participants compared to a Bayesian observer in gray, which represents normative performance on the task. Individual points are averages for each subject and filled in points represent group-level averages. Error bars are 95% confidence intervals. **B** Performance on the incremental learning task over time. Each timepoint represents ten trials. Lines are group averages, and bands are 95% confidence intervals. For normative comparison, the performance of the Bayesian observer is shown as a dotted gray line. **C** Performance on the incremental learning task as a function of task difficulty, which is indexed by the true underlying probability that pressing the F key was the correct response (>50%) on each trial. Points represent group level averages from 13 bins with an equal number of trials, lines represent the fit of a second-order linear model, and error bars and bands represent 95% confidence intervals. **D** Model performance of the Q Learner and baseline Biased Responder models. Left: Posterior predictive performance. Individual lines represent Q learner fits for each individual, whereas thick lines represent the group-level average fit (with the Q Learner in color and Biased Responder in gray). Bands represent 95% confidence intervals. Right: The difference in estimated out-of-sample predictive performance (as measured by expected log pointwise predictive density; ELPD) between the Q Learner and the Biased Responder model for each group. Individual points are the ELPD difference for each subject and filled in points represent group-level averages. Error bars are 95% confidence intervals

($\beta_{Group} = -0.88,\ 95\%\ CI = [-1.55,\ -0.144]$; Fig. 2A). CA participants' choices were less correct throughout the entirety of the task, even during periods of learning where action-outcome contingencies were more deterministic (e.g., close to 100%) compared to more difficult periods of learning ($\beta_{Group \times pFReward1^2} = -5.49,\ 95\%\ CI = [-7.57, -3.52]$; Fig. 2B–C. Overall, this difference in performance indicates that CA participants did not learn from reward feedback. Although CA participants responded slightly more slowly than healthy controls on this task ($\beta_{Group} = -115.81,\ 95\%\ CI = [-201.26, -33.55]$), we included reaction times as a covariate in the above regression analysis to ensure that differences in choice accuracy were not attributed to motor slowing in CA participants.

Next, to more formally assess participants' performance on this task, we fit a standard Q learning model to participants' responses. This model captures the extent to which each participant incorporated trial-by-trial outcomes into running estimates of the value of pressing each button in response to each cue, as well as whether choices are based on these estimates. As a baseline, we compared the performance of this model to a biased responder that merely estimated the extent to which each participant pressed one button over the other, regardless of outcome, in response to each cue. While healthy controls' responses were well described by the Q learning model, this model did no better than the biased responder at predicting CA participants' decisions, thus demonstrating that CA participants engaged in little-to-no incremental learning (Fig. 2D). On a measure of estimated out-of-sample predictive performance, controls were substantially better fit by the Q learner compared to the biased responder, while this improvement in fit was largely absent for CA participants ($\beta_{Group} = 30.94,\ 95\%\ CI = [16.465,\ 46.0]$). Thus, while healthy controls incorporated feedback into their estimates about the relationship between cue and action at each timepoint, CA participants generally did not.

Together, these results indicate that individuals with CA are impaired at reward learning from trial-and-error.

Impaired Incremental Reward Learning but Intact Episodic Memory in the Multiple Learning Strategies Task

After establishing that CA participants were impaired in a task that measured solely incremental reward learning, we wanted to examine both the specificity and generalizability of this impairment by (i) providing an alternative means of reward-based decision making alongside incremental learning and (ii) altering the incremental learning task structure to measure responses to reversal events rather than drifting probabilities. The multiple learning strategies task was thus used to accomplish both of these goals.

Consistent with the results of the incremental learning task, CA participants in the multiple learning strategies task were less responsive to reward outcomes compared to controls (Fig. 3A). Specifically, controls tended to choose the lucky deck more than CA participants immediately prior to a reversal ($\beta_{Group\times t=t-1} = 0.397,\ 95\%\ CI = [0.002,\ 0.807]$), and this tendency was disrupted by reversals; CA participants did not show this pattern ($\beta_{Group\times t=0} = -0.897,\ 95\%\ CI = [-1.28,\ -0.535]$) and remained below chance performance after a reversal occurred ($\beta_{Group\times t=t+1} = -0.595,\ 95\%\ CI = [-0.984,\ -0.21]$). This indicates that CA participants were unable to learn which deck had the higher expected value at any given time throughout the task. We next aimed to separately assess whether there were any differences between healthy controls and CA participants in the extent to which they chose previously seen (old) objects according to their value, which is a marker of using episodic memory to guide choices throughout the task. Although CA participants and healthy controls chose low-value, old objects at nearly identical rates ($\beta_{Group\times Value} = -0.220,\ 95\%\ CI = [0.031,\ -0.454]$), healthy controls chose high-value, old objects at a slightly higher rate than CA participants ($\beta_{Group\times Value} = 0.667,\ 95\%\ CI = [0.426,\ 0.904]$; Fig. 3B). This finding indicates that, relative to healthy controls, CA participants' choices were based primarily on episodic memory for low- rather than high-valued objects.

While these analyses assessed sensitivity to incrementally constructed value and episodic value separately, we next sought to capture the effects of each on participants' choices in a single model. To do so, we used a hybrid choice model, which combined a standard Q learning model with three inverse temperature parameters that captured each participants' sensitivity to estimated deck value (Deck Value), the true value of previously seen objects (Old Value), and a bias toward choosing previously seen objects regardless of their value (Old Bias; Fig. 3C). The first of these parameters measures the extent to which participants incorporated incrementally learned value into their choices, while the latter two measure the extent to which sources related to episodic memory impacted choice. For each group, we then assessed whether these inverse temperatures differed from zero and, if so, concluded that participants in that group made choices that were affected by each possible source. While healthy controls incorporated deck value into their decisions ($\beta_{HC} = 3.173,\ 95\%\ CI = [2.181,\ 4.189]$), CA participants generally did not ($\beta_{CA} = 0.681,\ 95\%\ CI = [-0.668,\ 2.066]$). This reward learning deficit was specific to value acquired incrementally, however, because CA participants and controls were both sensitive to episodic value, as measured by the value of old objects ($\beta_{HC} = 1.373,\ 95\%\ CI = [1.095,\ 1.654]$; $\beta_{CA} = 1.13,\ 95\%\ CI = [0.59,\ 1.643]$) and were both similarly biased by old objects regardless of their value ($\beta_{HC} = 1.142,\ 95\%\ CI = [0.798,\ 1.477]$; $\beta_{CA} = 0.551,\ 95\%\ CI = [0.028,\ 1.056]$). Furthermore, while there were no differences between groups for the effects of either episodic value ($\beta_{Group} = 0.244,\ 95\%\ CI = [-0.311,\ 0.820]$) or bias ($\beta_{Group} = 0.586,\ 95\%\ CI = [-0.051,\ 1.246]$), healthy controls were indeed more sensitive to learned deck value than CA participants ($\beta_{Group} = 2.47,\ 95\%\ CI = [0.362,\ 4.572]$). Finally, we compared the hybrid choice model to a biased responder, which again served as a baseline. The hybrid choice model outperformed this model, there was no difference between groups in estimated out-of-sample predictive performance ($\beta_{Group} = -1.631,\ 95\%\ CI = [-11.425,\ 8.444]$). Importantly, this indicates that the behavior of both CA participants and controls was well described by the hybrid choice model, which is expected if CA participants are unimpaired at episodic value learning.

We additionally had each participant complete a subsequent memory test for a subset of objects shown during the multiple learning strategies task. To capture

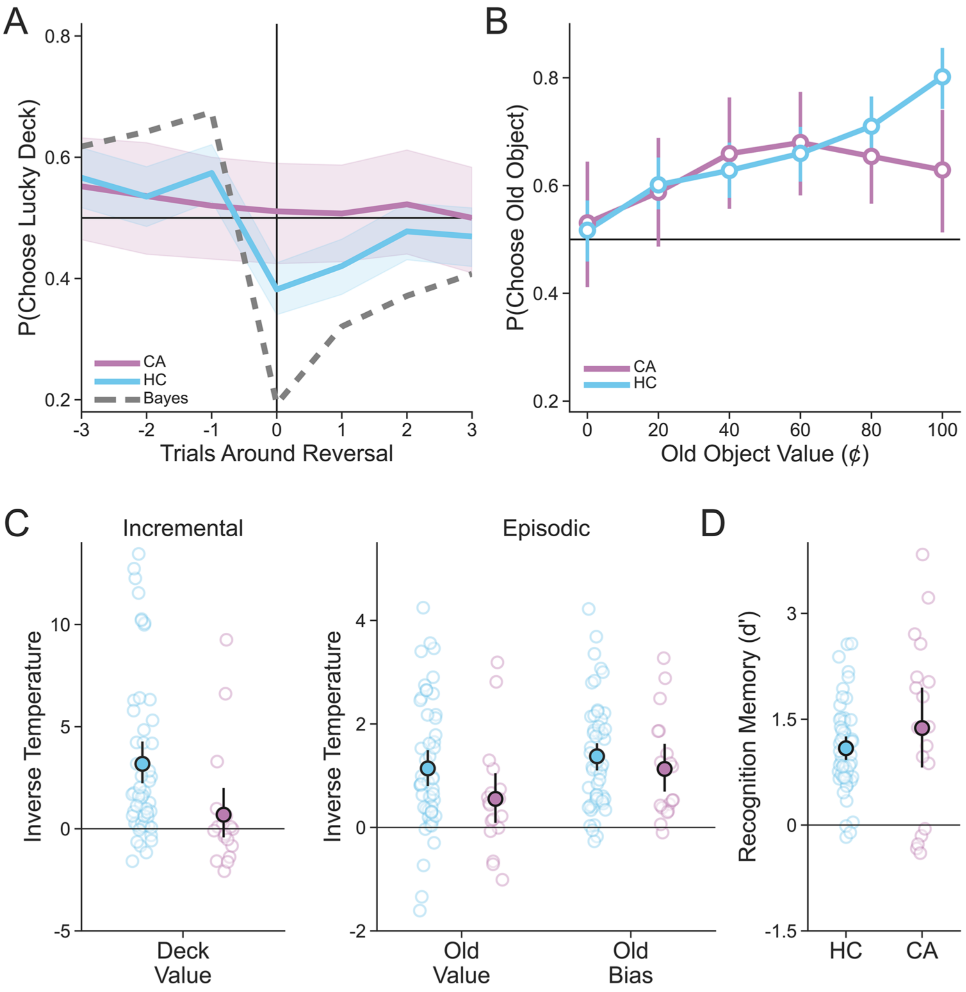

Fig. 3 Performance on the multiple learning strategies task. **A** Deck learning performance on the multiple learning strategies task as indicated by the proportion of trials on which the currently lucky deck was chosen as a function of how distant those trials were from a reversal in deck value. Performance for both healthy controls (HC) and CA participants is shown alongside a Bayesian observer with perfect episodic memory for visual comparison. Lines represent group averages, and bands represent 95% confidence intervals. **B** Object value usage on trials in which a previously seen object appeared. Points represent group averages and error bars represent 95% confidence intervals. **C** Inverse temperature estimates from the Hybrid model. Left: Inverse temperature for deck value (Deck Value), which captures impacts on choice related to incremental learning. Right: Inverse temperatures for old object value (Old Value) and a bias toward old objects regardless of their value (Old Bias), which capture impacts on choice related to episodic memory. Individual points represent estimates for each subject, group-level averages are shown as filled in points, and error bars represent 95% confidence intervals. **D** Recognition memory performance on the subsequent memory task. Individual points represent each participant's d-prime score, filled in points are group-level averages and error bars are 95% confidence intervals

performance on this task, we calculated the signal detection metric d-prime, which represents participants' ability to accurately discriminate between objects that were shown during the multiple learning strategies task and those that were brand new. There was no difference in recognition memory performance between groups ($\beta_{Group} = -0.487,\ 95\%\ CI = [-1.144,\ 0.157]$; Fig. 3D). This result provides further evidence that CA participants were unimpaired at using episodic memory throughout the task relative to their stark impairments in incremental learning. Lastly, CA participants and healthy controls demonstrated no differences in reaction time on this task ($\beta_{Group} = -86.40,\ 95\%\ CI = [-222.28, 44.76]$), suggesting that the behavioral differences reported here cannot be attributed to motor slowing in CA.

Controlling for Effects of Non-motor Deficits and Disease Subtype

We next sought to ensure that the differences in the tasks reported here were specific to deficits in reward learning rather than general cognitive impairment. Controlling for cognitive impairment is particularly important because recent work [43, 44] has suggested that incremental learning experiments tax higher level functions, like executive control and working memory, in addition to learning from reward prediction error. To address this issue and assess possible cognitive impairment, we conducted a battery of neuropsychological measures on CA participants (see Methods). Of these, a subset of measures were also completed by healthy controls (Supplementary Fig. 1). We found no differences in performance between groups on all measures except for the backwards digit span task, which indexes working memory ability, and on which healthy controls scored higher than CA participants (Table 2; $\beta_{Group} = -2.57,\ 95\%\ CI = [-4.18,\ -0.92]$). Backwards digit span scores were thus included as covariates in regression analyses where possible (see Methods) in order to control for impacts of this performance difference on impairments in incremental learning. To further ensure that CA participants' deficient incremental learning was not due to broad cognitive impairment, we also repeated all analyses excluding seven CA participants (and their matched controls) with mild cognitive impairment (MCI), as indicated by scoring lower than 26 on the MoCA (Table 1). While CA participants with MCI consisted of some of the lowest performing participants in our sample (Supplementary Figs. 2–3), there we no differences in the results across both tasks when they were excluded. It is therefore unlikely that CA participants' impaired reward learning ability is due to either working memory deficits or cognitive decline more broadly. A full report of these analyses can be found in Appendix 1.

Table 2 Neuropsychological test regression analysis results

Measure	β estimate	95% credible interval
Backwards Digit Span	– 2.57	[– 4.18, – 0.92]
Forwards Digit Span	– 0.09	[– 1.35, 1.15]
Semantic Fluency	1.54	[– 2.12, 5.14]
Phonemic Fluency	0.04	[– 2.75, 2.88]
Category Switching	1.68	[– 0.76, 4.10]
Similarities	– 0.20	[– 0.50, 0.12]
Go No Go	0.0001	[– 0.33, 0.33]

Results of regression analyses assessing differences in neuropsychological test performance between CA participants and healthy controls

We next sought to further characterize the nature of CA participants' reward learning impairment by looking at the relationship between incremental learning sensitivity, as measured by the Q learning models in each task, and performance on our neuropsychological battery. The extent to which CA participants learned about cues in the incremental learning task related only to total CCAS score ($r = 0.84,\ p < 0.001, \mathit{Bonferroni\ corrected}$; Table 3), suggesting that the specific contributions of the cerebellum to cognition may impact performance in this task. The CCAS scale was recently developed to measure the exact types of cognitive impairment that result from damage to the cerebellum [45]. Because more focal cerebellar lesions tend to lead to lower total CCAS scores [46], this provides further evidence of the necessity for the cerebellum to successfully perform the incremental learning task. The relationship between total CCAS score and performance was driven by the timed portions of the CCAS scale (e.g., the Semantic Fluency and

Table 3 Correlations between CA participant-level measures and incremental value sensitivity

Measure	Pearson's *R*	*p* value (Bonferroni corrected)	Task
Symptom duration	– 0.0661	1	Incremental learning
SARA (total)	0.1067	1	Incremental learning
MoCA (total)	0.5225	0.1885	Incremental learning
CCAS (Total)	0.8419	0.0001***	Incremental learning
BDI (total)	0.2977	1	Incremental learning
QUIP (total)	– 0.3892	0.7356	Incremental learning
Symptom duration	– 0.3699	0.7848	Multiple learning strategies
SARA (total)	– 0.2141	1	Multiple learning strategies
MOCA (total)	0.1438	1	Multiple learning strategies
CCAS (total)	0.3849	0.6887	Multiple learning strategies
BDI (total)	0.4641	0.3141	Multiple learning strategies
QUIP (total)	– 0.074	1	Multiple learning strategies

CA participant-level measures consist of total neuropsychological scores and symptom duration, and incremental value sensitivity consists of estimates by the Q Learning model in the incremental learning task and as by the Hybrid Q Learning model in the multiple learning strategies task

CCAS cerebellar cognitive affective/Schmahmann syndrome scale, *SARA* Scale for the Assessment and Rating of Ataxia, *MoCA* Montreal cognitive assessment, *BDI* Beck's depression inventory, *QUIP* questionnaire for impulsive-compulsive disorders in Parkinson's disease

$^{***}p < 0.001$

Table 4 Correlations between CA participant CCAS subscale measures and incremental value sensitivity

CCAS measure	Pearson's *R*	*p* value (Bonferroni corrected)
Semantic Fluency	0.6693	0.033*
Phonetic Fluency	0.6235	0.0749
Category Switching	0.7168	0.012*
Digit Span Fwd (CCAS)	0.295	1
Digit Span Bwd (CCAS)	0.3146	1
Cube Drawing	0.5009	0.4055
Verbal Recall	0.4224	0.9118
Similarities	0.5881	0.1302
Go No Go	– 0.2654	1
Affect	0.2944	1

Incremental value sensitivity consists of estimates by the Q Learning model in the incremental learning task

*$p < 0.05$

Category Switching measures; Table 4), suggesting a potential effect of slowed responses in the incremental learning task. While CA participants did indeed respond more slowly than healthy controls on this task (see above), we controlled for this difference in our behavioral analysis. Finally, there was no relationship between any measure and incremental learning ability in the multiple learning strategies task (Table 3).

Finally, we addressed the possibility that the subset of our sample of CA participants consisting of diagnoses that were less restricted to the cerebellum, namely the three individuals with multiple system atrophy (MSA) and the two individuals with Friedrich's ataxia (FA), could be responsible for the deficits reported here. We repeated all analyses with these five participants excluded and found no differences in the results (Supplementary Figs. 4–5). A full report of these analyses can be found in Appendix 2.

Discussion

The results of the present work demonstrate that individuals with cerebellar dysfunction, represented by CA cases in our cohort, are impaired at trial-and-error reward learning. While the cerebellum and basal ganglia have traditionally been treated as making separate contributions to learning [5, 21], recent findings have called this dichotomy into question [8–19]. This work has suggested that, alongside its role in motor learning, the cerebellum likely operates in concert with the basal ganglia to support reinforcement learning from reward. Our study corroborates these findings from animal models [10–19], providing evidence that the human cerebellum is necessary for learning associations from reward. In comparison to age- and sex-matched healthy controls, CA participants were impaired at reward-based learning from trial-and-error. Further, CA participants retained the ability to employ an alternative strategy based in episodic memory to guide their decisions, demonstrating that this impairment is specific to incremental learning. These results challenge the idea that the cerebellum is used primarily for motor learning and shed light on how multiple neural systems may interact with one another to support learning in the non-motor domain.

Our findings join a litany of recent research suggesting that the cerebellum plays a broad role in human cognition [22, 47–49]. Indeed, individuals with damage to the cerebellum demonstrate impairment in a wide range of cognitive functions including cognitive control [50] and impulsivity [51]. Human functional neuroimaging studies have also revealed cerebellar activity in a variety of different non-motor tasks [22, 48]. Many of these functions are likely supported by the robust bidirectional connections the cerebellum shares with the prefrontal cortex [52, 53]. In particular, recent findings have indicated that individuals with CA have heightened domain-specific impulsive and compulsive behaviors, which is a common symptom of underlying reward system dysfunction [54, 55]. Our study adds to this work by suggesting that the cerebellum is additionally necessary for reward learning in humans.

While there is growing evidence validating the implication of the cerebellum in reward-based learning in animals, there is only limited work on this topic in humans. Early imaging studies, for example, demonstrated cerebellar BOLD activity in patients with substance use disorder who performed reward-based learning tasks [23] and experienced cravings [24], and also in response to unexpected reward [25]. However, it remains unknown how cerebellar damage impacts reward learning, as investigations of reward learning in the cerebellum are rare. While two previous studies employed reward-based experimental tasks in individuals with isolated ischemic lesions of the cerebellum [56, 57], results until this point have remained far from conclusive. Thoma et al. (2008) used a reward-based learning task consisting of an initial acquisition phase in which eight participants with cerebellar damage were rewarded for learning associations between colors and symbols followed by a reversal portion in which they had to disremember previously acquired knowledge and learn new associations for each cue. While participants with cerebellar damage demonstrated no impairment at acquiring new, reward-based knowledge, they were selectively impaired at learning from a single reversal. While this study complements our findings, we found evidence for more global impairment: CA participants in both of our tasks were unable to learn associations from reward on a trial-by-trial basis. Rustemeier et al. (2016) took a different approach by asking twelve individuals with cerebellar damage to learn a simple acquisition task from probabilistic feedback and subsequently transfer this knowledge to re-arranged stimuli. While participants were unimpaired behaviorally at this task, electroencephalographic

(EEG) results revealed that they may process reward-based feedback differently from controls. Our findings support this interpretation and further suggest that processing of trial-by-trial feedback is not just different, but impaired, in individuals with cerebellar damage. Finally, while other related studies showed impairment in learning from reinforcement in participants with cerebellar damage [20, 58], this work has focused primarily on movement-dependent deficiencies.

While our findings suggest that the cerebellum is necessary for incremental reward learning, they cannot speak to the neural circuitry underlying this role. One intriguing possibility is that the cerebellum may operate in tandem with the basal ganglia—canonically seen as the seat of reinforcement learning in the brain [5, 21]—to learn about reward incrementally. Reward prediction error signals in midbrain dopamine neurons that provide input to the basal ganglia [27, 29] have also been found to be encoded by cerebellar neurons [9, 15, 17, 19]. Further, through excitatory projections to the ventral tegmental area, the cerebellum has widespread reciprocal connections with the basal ganglia and has recently been shown to influence reward-driven behavior through these projections [8, 59]. While reinforcement learning via the basal ganglia and supervised learning via the cerebellum have typically been treated as fulfilling entirely separate roles [5, 21], these systems appear to be more interdependent than previously thought. Future investigations of the relationship between the basal ganglia and cerebellum are needed to clarify the exact mechanisms underlying reinforcement learning in the brain.

Lastly, there are several potential limitations related to the nature of our sample that should be considered when interpreting these findings. First, cerebellar dysfunction in our sample of CA participants was caused by several different conditions. While most of these pathologies are predominantly restricted to the cerebellum, non-cerebellar brain areas and circuits could also be affected, particularly in participants diagnosed with either MSA or FA. There was, however, no change in the reported reward-based learning deficits when these participants were excluded. Second, while cognitive impairment due to neurodegenerative disease could potentially contribute to some of the deficits measured here, we accounted for this possibility by establishing that the incremental reward learning deficits reported here persist regardless of MCI status. We also collected basic neuropsychological measures from all participants, and CA participants were not different from controls on the vast majority of measures. We focused particularly on possible contributions of working memory given recent work suggesting that working memory plays an important role in incremental reward learning [43, 44]. While CA participants and controls performed similarly on the forward digit span task, CA participants were somewhat impaired at backwards digit span. We controlled for this difference by including backwards digit span scores as covariates in our analyses. Finally, while our control participants completed the study online, we accounted for potential variability caused by this difference in setting by collecting three matched controls for each CA participant in our sample.

Taken together, our findings suggest that the human cerebellum is necessary for reward learning. These results provide new constraints on models of non-motor learning and suggest that the cerebellum and basal ganglia work in tandem to support learning from reinforcement.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12311-023-01633-2.

Author Contribution J.N., C.R.L., L.M.K., M.K.P., S.K., and D.S. designed the study. C.A., N.D., and C.R.L. collected data from cerebellar ataxia participants. J.N. collected data from healthy control participants. J.N. analyzed the data and prepared all figures. J.N. and C.A. wrote the main manuscript text. All authors reviewed the manuscript.

Funding J.N. was supported by the NSF Graduate Research Fellowship (1644869). S.H.K. was supported by NINDS R01NS104423, NINDS R01 NS118179, NINDS R01 NS124854, and National Ataxia Foundation. D.S. was supported by an NSF CRCNS award (1822619), NIMH R01 MH121093 and the Kavli Foundation.

Data Availability All code used to analyze the data in this study may be found here: https://github.com/boomsbloom/ataxia-rl. Data is available upon request from the corresponding authors.

Declarations

Ethical Approval Informed consent from all participants in the study was obtained with approval from the Columbia University Institutional Review Board.

Competing Interests The authors declare no competing interests.

References

1. Raymond JL, Lisberger SG, Mauk MD. The cerebellum: a neuronal learning machine? Science. 1996;272:1126–31.
2. Llinás R, Welsh JP. On the cerebellum and motor learning. Curr Opin Neurobiol. 1993;3:958–65.
3. Ito M, Itō M. The cerebellum and neural control. 1984;(Raven Press)
4. Marr D. A theory of cerebellar cortex. J Physiol. 1969;202:437–70.
5. Doya K. What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Netw. 1999;12:961–74.
6. Wolpert DM, Miall RC, Kawato M. Internal models in the cerebellum. Trends Cogn Sci. 1998;2:338–47.
7. Raymond JL, Medina JF. Computational principles of supervised learning in the cerebellum. Annu Rev Neurosci. 2018;41:233–53.
8. Caligiore D, Arbib MA, Miall RC, Baldassarre G. The super-learning hypothesis: integrating learning processes across cortex, cerebellum and basal ganglia. Neurosci Biobehav Rev. 2019;100:19–34.
9. Hull C. Prediction signals in the cerebellum: beyond supervised motor learning. eLife. 2020;9:e54073.
10. Sendhilnathan N, Goldberg ME. The mid-lateral cerebellum is necessary for reinforcement learning. 2020. http://biorxiv.org/lookup/doi/10.1101/2020.03.20.000190

11. Sendhilnathan N, Semework M, Goldberg ME, Ipata AE. Neural correlates of reinforcement learning in mid-lateral cerebellum. Neuron. 2020;106:188-198.e5.
12. Sendhilnathan N, Ipata A, Goldberg ME. Mid-lateral cerebellar complex spikes encode multiple independent reward-related signals during reinforcement learning. Nat Commun. 2021;12:6475.
13 Larry N, Yarkoni M, Lixenberg A, Joshua M. Cerebellar climbing fibers encode expected reward size. eLife. 2019;8:e46870.
14 Carta I, Chen CH, Schott AL, Dorizan S, Khodakhah K. Cerebellar modulation of the reward circuitry and social behavior. Science. 2019;363:eaav0581.
15 Heffley W, Hull C. Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. eLife. 2019;8:e46764.
16. Wagner MJ, Kim TH, Savall J, Schnitzer MJ, Luo L. Cerebellar granule cells encode the expectation of reward. Nature. 2017;544:96–100.
17. Heffley W, et al. Coordinated cerebellar climbing fiber activity signals learned sensorimotor predictions. Nat Neurosci. 2018;21:1431–41.
18. Kostadinov D, Beau M, Blanco-Pozo M, Häusser M. Predictive and reactive reward signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. Nat Neurosci. 2019;22:950–62.
19. Ohmae S, Medina JF. Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. Nat Neurosci. 2015;18:1798–803.
20. Therrien AS, Wolpert DM, Bastian AJ. Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. Brain J Neurol. 2016;139:101–14.
21. Doya K. Complementary roles of basal ganglia and cerebellum in learning and motor control. Curr Opin Neurobiol. 2000;10:732–9.
22. King M, Hernandez-Castillo CR, Poldrack RA, Ivry RB, Diedrichsen J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. Nat Neurosci. 2019;22:1371–8.
23. Volkow ND, et al. Expectation enhances the regional brain metabolic and the reinforcing effects of stimulants in cocaine abusers. J Neurosci Off J Soc Neurosci. 2003;23:11461–8.
24. Grant S, et al. Activation of memory circuits during cue-elicited cocaine craving. Proc Natl Acad Sci U S A. 1996;93:12040–5.
25. Ramnani N, Elliott R, Athwal BS, Passingham RE. Prediction error for free monetary reward in the human prefrontal cortex. Neuroimage. 2004;23:777–86.
26. Sutton RS, Barto AG. Reinforcement learning: an introduction. 352.
27. Houk JC, Adams JL, Barto AG. A model of how the basal ganglia generate and use neural signals that predict reinforcement. in Models of information processing in the basal ganglia 249–270 (The MIT Press, 1995).
28. Rescorla RA, Wagner AR. 3 A theory of Pavlovian conditioning : variations in the effectiveness of reinforcement and nonreinforcement. in 1972
29. Schultz W, Dayan P, Montague PR. A Neural substrate of prediction and reward. Science. 1997;275:1593–9.
30. Kuo S-H. Ataxia. Contin Minneap Minn. 2019;25:1036–54.
31. Duncan K, Semmler A, Shohamy D. Modulating the use of multiple memory systems in value-based decisions with contextual novelty. J Cogn Neurosci. 2019;1–13. https://doi.org/10.1162/jocn_a_01447
32 Nicholas J, Daw ND, Shohamy D. Uncertainty alters the balance between incremental learning and episodic memory. eLife. 2022;11:e81679.
33. Hariri AR. The emerging importance of the cerebellum in broad risk for psychopathology. Neuron. 2019;102:17–20.
34. Bellebaum C, Daum I. Cerebellar involvement in executive control. Cerebellum. 2007;6:184–92.
35. Beuriat P-A et al. A new insight on the role of the cerebellum for executive functions and emotion processing in adults. Front Neurol. 2020;11
36. Mannarelli D, et al. The cerebellum modulates attention network functioning: evidence from a cerebellar transcranial direct current stimulation and attention network test study. Cerebellum. 2019;18:457–68.
37. Litman L, Robinson J, Abberbock T. TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. Behav Res Methods. 2017;49:433–42.
38. Hoffman MD, Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. 31.
39. Team SD. Stan Reference Manual.
40. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017;27:1413–32.
41. Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng. 1960;82:35–45.
42. Nassar MR, Wilson RC, Heasly B, Gold JI. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J Neurosci. 2010;30:12366–78.
43. Collins AGE, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci. 2012;35:1024–35.
44. Yoo AH, Collins AGE. How working memory and reinforcement learning are intertwined: a cognitive, neural, and computational perspective. J Cogn Neurosci. 2022;34:551–68.
45. Hoche F, Guell X, Vangel MG, Sherman JC, Schmahmann JD. The cerebellar cognitive affective/Schmahmann syndrome scale. Brain. 2018;141:248–70.
46. Chirino-Pérez A, et al. Mapping the cerebellar cognitive affective syndrome in patients with chronic cerebellar strokes. Cerebellum. 2022;21:208–18.
47. McDougle SD et al. Continuous manipulation of mental representations is compromised in cerebellar degeneration. Brain J Neurol. 2022;awac072. https://doi.org/10.1093/brain/awac072.
48. Buckner RL. The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. Neuron. 2013;80:807–15.
49. Koziol LF, et al. Consensus paper: The cerebellum's role in movement and cognition. Cerebellum Lond Engl. 2014;13:151–77.
50. Alexander MP, Gillingham S, Schweizer T, Stuss DT. Cognitive impairments due to focal cerebellar injuries in adults. Cortex J Devoted Study Nerv Syst Behav. 2012;48:980–90.
51. Amokrane N, Lin C-YR, Desai NA, Kuo S-H. The impact of compulsivity and impulsivity in cerebellar ataxia: a case series. Tremor Hyperkinetic Mov. 10;43
52. Buckner RL, Krienen FM, Castellanos A, Diaz JC, Yeo BTT. The organization of the human cerebellum estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106:2322–45.
53. Middleton FA, Strick PL. Cerebellar projections to the prefrontal cortex of the primate. J Neurosci. 2001;21:700–12.
54. Amokrane N, et al. Impulsivity in cerebellar ataxias: testing the cerebellar reward hypothesis in humans. Mov Disord. 2020;35:1491–3.
55. Chen TX, et al. Impulsivity trait profiles in patients with cerebellar ataxia and Parkinson disease. Neurology. 2022;99:e176–86.
56. Thoma P, Bellebaum C, Koch B, Schwarz M, Daum I. The cerebellum is involved in reward-based reversal learning. Cerebellum. 2008;7:433.
57. Rustemeier M, Koch B, Schwarz M, Bellebaum C. Processing of positive and negative feedback in patients with cerebellar lesions. Cerebellum Lond Engl. 2016;15:425–38.
58. McDougle SD, et al. Credit assignment in movement-dependent reinforcement learning. Proc Natl Acad Sci. 2016;113:6797–802.
59. Caligiore D, et al. Consensus paper: Towards a systems-level view of cerebellar function: the interplay between cerebellum, basal ganglia, and cortex. Cerebellum. 2017;16:203–29.

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