Reinforcement motor learning after cerebellar damage is related to state estimation.

Christopher White[1], Evan C. Snow[1], Amanda S. Therrien[1,2]

1. Moss Rehabilitation Research Institute, Elkins Park, PA

2. Department of Rehabilitation Medicine, Thomas Jefferson University, Philadelphia, PA

Please address correspondence to:

Amanda S. Therrien, PhD

Moss Rehabilitation Research Institute

Medical Arts Building, Suite 100

50 Township Line Rd. Elkins Park, PA, 19027

Email: Amanda.Therrien@jefferson.edu

ABSTRACT

Growing evidence suggests that the cerebellum may play a role in neural mechanisms of reinforcement learning, but the nature of its contribution remains unclear. Recent work posited that cerebellum-dependent sensory prediction contributes to reinforcement learning in motor contexts by enhancing body state estimates, which are necessary to solve the credit-assignment problem. The objective of this study was to test the relationship between the predictive component of state estimation and reinforcement motor learning in individuals with cerebellar degeneration. Individuals with cerebellar degeneration and neurotypical control participants completed two tasks: a reinforcement learning task that required them to alter the angle of reaching movements and a state estimation task that tested the somatosensory perception of active and passive movement. The state estimation task permitted calculation of the active benefit shown by each participant, which is thought to reflect the cerebellum-dependent predictive component of state estimation. We found that the cerebellar and control groups showed similar magnitudes of learning with reinforcement and active benefit on average, but there was substantial variability across individuals. Using multiple regression, we assessed potential predictors of reinforcement learning within the cerebellar group. Our analysis included active benefit, clinical ataxia severity, movement variability, movement speed, and age as independent variables. We found a significant relationship in which greater active benefit predicted better learning with reinforcement. No other variables showed significant relationships with learning. Overall, our results support models that posit an indirect role for the cerebellum in reinforcement motor learning: one in which it may contribute to state estimation.

KEYWORDS

Cerebellum, degeneration, internal model, reaching, somatosensation

INTRODUCTION

The cerebellum has long been known to play a critical role in controlling movement [1]. A dominant hypothesis is that the cerebellum houses internal models of body and environment dynamics, which it combines with information about motor commands to generate predictions of the sensory consequences of actions (i.e., the forward internal model) [2,3]. Sensory predictions serve as internal feedback signals that compensate for time-delayed peripheral afference, permitting rapid and well-coordinated movement. In the event of changes to the body or environment dynamics, a form of supervised motor learning, often called adaptation, is thought to recalibrate internal models (and sensory predictions) to keep them up to date [4,5]. A large body of literature supports this hypothesis, demonstrating how the motor symptoms of cerebellar damage correspond with disrupted internal model and sensory prediction calculation, as well as how cerebellar damage impairs adaptation (for work in humans, see [6–8] and [9–16], respectively).

Motor learning is achieved through a combination of multiple neural mechanisms [17,18], and the cerebellar contribution to motor learning mechanisms other than adaption is less well understood. In recent years, the potential role of the cerebellum in reinforcement motor learning (RL) has received particular attention [19–21]. RL is a form of semi-supervised learning. While adaptation relies on continuous feedback about the direction and magnitude of movement errors, RL relies on periodic scalar measures of outcome [22,23]. At the simplest level, reinforcement feedback can take the form of binary signals indicating success or failure. RL requires individuals to explore possible task solutions to determine which one(s) yield(s) success. Over

time, the probability of success is learned for each task solution, and this permits a biasing of motor selection to favor the solution(s) most likely to generate a positive outcome.

Therrien et al. (2016) [19] showed that individuals with cerebellar degeneration could use binary reinforcement to learn to rotate the angle of a reaching movement. However, the cerebellar group learned more slowly than neurotypical control participants on average, and there was marked variability in the degree of learning observed across individuals. A mathematical model of the RL task posited that efficient learning from binary reinforcement required accurate state estimation. In other words, to solve the credit assignment problem and map reinforcement signals to the correct movement, the motor system must know what action was produced. The model parsed trial-to-trial variability in reach angle as random draws from two independent distributions: exploration variability and sensorimotor noise. It assumed that the proportion of variability from exploration could be estimated, but the proportion of variability from sensorimotor noise could not. When a movement was reinforced, learning could only be based on the proportion of total variability that could be estimated. Fitting the model to learning data showed that the cerebellar group's behavior was consistent with a greater proportion of trial-to-trial variability stemming from the sensorimotor noise distribution than neurotypical controls. This result suggested that the cerebellar group learned less efficiently due to greater uncertainty in state estimation. While RL was intact overall after cerebellar damage, it seemed the cerebellum might play an indirect role in RL when body state estimation is required. Indeed, others have proposed a similar idea using different models [20,24].

Sensory prediction is thought to contribute to state estimation [25,26]. The contribution is evidenced by an enhancement of somatosensory acuity during active (i.e., voluntary) compared to passively induced movement [27–29]. Cerebellar degeneration has been shown to disrupt this 'active benefit' [30–33]; however, Weeks et al. (2017a) [32] demonstrated that the degree of disruption can vary across individuals. It is conceivable, then, that the magnitude of active benefit shown by an individual with cerebellar damage could constitute a metric of the integrity of the forward internal model contribution to state estimation. A correlation between active benefit and RL in individuals with cerebellar damage would support the central hypothesis of the model proposed by Therrien et al. (2016) [19].

Impaired predictive state estimation is a parsimonious explanation for inefficient motor RL after cerebellar damage, yet it is far from the only possibility. The severity of individuals' ataxia, or their baseline movement variability, may impede the performance of the correct task solution regardless of whether an individual can discern what that is. Simply reducing movement velocity can additionally improve motor output in this population (it makes the movement dynamics easier to control and increases the time available to use delayed peripheral feedback) and, hence, permit better expression of learning [6,7,34]. Finally, learning with binary feedback has been shown to rely heavily on explicit strategy use in younger neurotypical populations [35,36]. Cerebellar damage has been implicated in the learning of explicit strategies (vs. their execution) [37,38], although there is evidence that participant age may pose a larger constraint on strategy computation [39]. Aging is also associated with reduced sensitivity to feedback in non-motor reward learning tasks [40–42]. Consequently, variation in RL across individuals could reflect age-related declines in cognitive capacity.

Here, we compared individuals with cerebellar degeneration to a group of neurotypical control participants matched for age and biological sex as they performed an RL task that required them to alter their reach angle. Both groups also performed a somatosensory state estimation task that tested the sensory perception of active and passive movement. Within the cerebellar group, we tested the relationship between the magnitude of learning in the RL task and the degree of active benefit shown in the state estimation task. Importantly, we used multiple regression to test the significance of this relationship in the presence of other possible predictors of learning: ataxia severity, baseline movement variability, change in reach velocity, and participant age. Our results showed that the cerebellar and control groups performed similarly in both tasks, on average, but there was substantial inter-individual variability. In the cerebellar group, the magnitude of learning in the RL task showed a significant relationship with the degree of active benefit in the state estimation task. No other predictors showed significant relationships with learning. We discuss the significance of these findings by considering how they contribute to an understanding of a cerebellar role in motor RL and the potential application of RL in rehabilitation training protocols for individuals with cerebellar damage.

MATERIALS AND METHODS

Participants

The study was approved by the Einstein Healthcare Network and Jefferson Health Institutional Review Boards, and all participants gave written informed consent before participation. A total of 12 people with cerebellar degeneration (CB; 5 female, 7 male, mean age 55.1 years, SD: 15.6

years) and 12 neurotypical individuals (Control; 5 female, 7 male, mean age 55.8 years, SD: 15.8 years) participated in the study. Further details about the CB group's characteristics are shown in Table 1. CB participants were screened for and presented with no extra cerebellar or extrapyramidal signs on a neurological exam, normal passive peripheral sensation in the upper extremities, and no history of drug or alcohol abuse or other neurological condition. One participant with cerebellar degeneration (CB05) presented with hyperreflexia in the lower limbs; however, reflexes in the upper limbs were within normal limits. Control participants were also screened for and presented with no history of drug or alcohol abuse or neurological condition. CB participants' movement impairment severity was assessed using the International Cooperative Ataxia Rating Scale (ICARS) [43].

Apparatus

The experimental tasks were performed using the KINARM Exoskeleton robot (Kinarm, Kingston, Ontario, Canada). Each participant's dominant arm rested in trays in this device and performed reaching movements in the horizontal plane below a screen that prevented them from directly viewing their arm. Throughout the experimental tasks, visual feedback was projected onto the screen surface.

Procedure

Participants completed two tasks with their self-reported dominant arm: a reinforcement learning task and a somatosensory state estimation task. The tasks were performed in separate sessions, and the order was fixed so that participants completed the reinforcement motor learning task first.

Reinforcement Learning (RL) Task

Figure 1a shows a schematic of the task. This task was adapted from Therrien et al. 2016 [19] and required participants to learn to rotate the angle of 10 cm, target-directed reaching movements in a counterclockwise direction by 15 degrees. The reach angle was defined as the angle between two vectors: one from the center of the home position to the center of the target and another from the center of the home position to the index fingertip location when it crossed a radial distance of 10 cm from the center of the home position (i.e., the reach endpoint). Before performing the task, participants were given time to practice making the target-directed reaching movements with and without visual feedback of their arm to familiarize themselves with the Kinarm and task environment. In the experimental task, all participants performed a 70-trial baseline phase with no rotation. Visual feedback of the hand position was provided for the first 10 trials of the baseline phase and then removed for the remaining 60 trials. Only data from the 60 baseline trials without visual feedback was used for analysis. Following the baseline phase, participants completed a 180-trial perturbation phase in which participants learned to rotate the reach endpoint. The perturbation phase here was deliberately shorter than the same phase in Therrien et al. (2016) [19] to focus on the early learning trials in which the rate difference between the cerebellar and control group was most apparent in the prior study. Participants were given a short break (1-3 min) every 60 trials throughout the task to avoid the effects of fatigue.

A trial began with participants' index fingertips in a home position (1 cm radius circle) located 35 cm in front of the chest and aligned to the sagittal midline. A cursor (0.5 cm circle) was projected over the index fingertip. Once the arm was held in that position for 500 ms, the cursor

was extinguished, and a target appeared on the screen (1 cm radius circle). The center of the target was positioned 10 cm distal to the home position, directly in front of it. Participants were instructed to begin their reach when they saw the target appear and to reach so that the index fingertip passed through the center of the target. The trial ended when the index fingertip exceeded a radial distance of 10 cm from the center of the home position. The index fingertip position at this point was used to calculate the reach endpoint angle. After a post-trial delay of 700 ms, the target was extinguished, and the participant could move their hand toward their chest. Once the index fingertip was within a 6 cm radius of the home position, the fingertip cursor reappeared to help participants reposition the arm for the next trial. To encourage participants to complete their movement within 200-600 ms after leaving the home position, the text "Too Fast" or "Too Slow" was displayed on the screen in the post-trial delay period following reaches that fell outside this range.

Throughout the task, participants received binary (i.e., success or failure) feedback about the outcome of their reaches. Reaches in which the index fingertip crossed the reward zone were reinforced with the target changing color from light blue to green and an auditory tone (Figure 1a). Reaches where the index fingertip missed the reward zone resulted in the target changing color from light blue to red and no auditory tone. Participants were not informed about the component of their reaches mapped to the feedback signals. Rather, they were instructed to use the binary feedback provided to discern the correct reaching movement throughout the task. During the baseline phase, the reward zone corresponded to the visually presented target – i.e., the index finger had to pass through the visually displayed target. During the perturbation phase, a closed-loop reinforcement schedule (Therrien et al. 2016 [19], Figure 1b) was used to drive

participants to rotate the endpoint of the reach in a counterclockwise direction. The reward zone was determined by calculating a moving average of the endpoint angles from the previous 10 trials. The reward zone corresponded to the angles between the moving average and the rotated target.

Somatosensory State Estimation Task

This task was adapted from Weeks et al. 2017b [33]. It required participants to report the location of a visual cursor relative to the perceived position of the unseen index finger in a 2-alternative forced choice paradigm after reaches were performed voluntarily (active movement condition) or were passively induced by the robot (passive movement condition). Reaches were performed with the dominant arm along the y-axis of the workspace to make them comparable to the movements performed in the RL task. Somatosensory perception was assayed at the reach endpoint because the RL task required participants to learn to bias the position of the reach endpoint.

Figure 1c shows a schematic of the task. All trials began with the robot passively moving the arm to place the index fingertip in a home position (1 cm radius circle). A visual cursor then displayed the index fingertip position for 2 s before being extinguished along with the home position. In the passive movement condition, the robot moved the hand to one of five possible probe locations along a 20 cm reach path. In the active movement condition, an auditory "go" signal cued participants to reach voluntarily until they encountered a haptic wall that stopped the reach. The five probe locations prevented participants from habituating to a single arm movement. At the probe location, the visual cursor reappeared for a 600 ms display period, in

which its position was shifted relative to the index fingertip (±100, ±80, ±60, ±40, ±30, ±20, ±13, ±7, 0 mm). Following the display period, the cursor was extinguished. Depending on the movement condition, participants completed the remainder of the 20 cm reach, either passively or actively. In the active movement condition, the reach endpoint was signaled by a second haptic wall. At the reach endpoint, participants verbally reported whether they perceived the cursor had been "ahead" of or "behind" their index fingertip during the display period. The experimenter recorded the participants' reports, and then the robot passively moved the arm back to the home position to begin the next trial.

In the active movement condition, a 5 cm wide force channel was used to constrain reach trajectories to the desired region of the workspace. The width of the channel ensured that participants could not rely entirely on the haptic walls to guide their reach and perform the task correctly. This prevented them from using the channel wall as an additional source of sensory information in the task. Participants were moved at 17 cm/s in the passive movement condition, corresponding to the slowest permittable movement speed in the reinforcement motor learning task. Participants were encouraged to keep their average movement speed between 7 – 27 cm/s in the active movement condition. The text "Too Fast" or "Too Slow" was displayed on the screen at the end of trials when movement speeds fell outside this range.

The probe location and shift magnitude were randomized across trials in both movement conditions. Using the method of constant stimuli, participants completed 10 repetitions of each shift magnitude, corresponding to 2 repetitions of each shift magnitude per probe location and a

total of 170 trials per movement condition. The order of passive and active movement conditions was counterbalanced across participants.

Analysis

All kinematic analysis was performed using custom-written scripts in Matlab (Mathworks), and statistical analysis was performed using SPSS v29 (IBM). Experimental data and analysis scripts are available for download here:
https://osf.io/s86bp/?view_only=c85f779a6f734825be1950dd54145fe0

In the RL task, reach angles were computed from the raw position data and expressed relative to the target location so that angles in the clockwise direction were positive and angles in the counterclockwise direction were negative. The dependent variables of mean reach angle, reach angle standard deviation, and reach peak velocity were assessed using mixed analyses of variance (ANOVA). Between-subjects comparisons were made across a factor of group (cerebellar degeneration, CB; neurotypical, Control), and within-subjects comparisons were made over two experimental phases: the final 20 trials of the baseline phase and the final 20 trials of the perturbation phase. Total learning – i.e., the difference between the end of the perturbation and baseline phases – was compared across groups using an independent samples t-test. Within each group, the total learning was compared against zero using one-sample t-tests. Finally, the change in peak velocity between the baseline and the end of the perturbation phases was compared across groups using an independent samples t-test and compared to zero within groups using one-sample tests. Change values below zero indicated that participants increased

their speed from baseline to the end of the perturbation phase, whereas positive values indicated that they slowed down.

For the somatosensory state estimation task, the proportion of trials in which a participant reported the cursor to be "ahead" of the index fingertip was determined for each shift value. The data were then fit with a psychometric (logistic) function [44]. Separate functions were fit for each subject and movement condition. Three dependent variables were derived from the fitted psychometric functions: the just noticeable difference (JND), the active benefit, and the bias. The JND quantified the slope of the psychometric function and served as a measure of somatosensory acuity. The JND was calculated for each subject and movement condition as the difference in shift values (i.e., the x-axis in Figure 1d) corresponding to the 75% and 25% probabilities (i.e., the y-axis in Figure 1d) on the psychometric function. A smaller JND indicated better somatosensory acuity. We calculated active benefit as the difference in JND between the active and passive movement conditions. Values below zero indicated active benefit (i.e., the active JND was smaller than the passive JND). The bias quantified the midpoint of the psychometric function and served as a measure of the point of subjective equality (i.e., the shift value at which participants perceived the cursor to be aligned with the index fingertip position). The bias was calculated as the shift value corresponding to a 50% probability on the psychometric function. The statistical analyses of JND and bias used mixed ANOVA with a between-subjects factor of group (CB, Control) and a within-subjects factor of movement condition (passive, active). The active benefit was compared across groups using an independent samples t-test and compared against zero within groups using one-sample t-tests.

Within the active movement condition, additional analyses were performed to compare movement speed and trajectory variability across groups. Specifically, independent samples t-tests were used to compare the reach peak velocity and the lateral variability of reach movements (i.e., the standard deviation of movement along the x-axis) between the Control and CB groups. We also assessed the bivariate correlations (Pearson r) between these measures and the active benefit within each group.

Finally, we used multiple linear regression analysis to assess predictors of reinforcement motor learning in the CB group. Total learning in the RL task was the dependent variable. The following independent variables were set as predictors in the regression model: the limb kinetic sub-score of the ICARS (i.e., the severity of CB participants' limb ataxia), the baseline reach angle variability in the RL task, the change in peak reach velocity in the RL task, participants' age, and the active benefit shown in the somatosensory state estimation task.

All data were tested for normality using Shapiro-Wilk tests. Homogeneity of variance was assessed using Mauchly's Test of Sphericity. To ensure that data met the assumptions for multiple regression analysis, we used the Durbin-Watson test for independence of observations, assessed the Tolerance/VIF values for multicollinearity, and the normal P-P plot for approximately normally distributed residuals. For all statistical analyses, the alpha value was set to 0.05.

RESULTS

Figure 2A shows the time series of reach angles for the cerebellar degeneration (CB) and neurotypical (Control) groups in the reinforcement learning (RL) task. Analysis of variance (ANOVA) was used to compare the mean reach angle between groups and across two task phases (the final 20 trials baseline and the final 20 trials of the perturbation phases). This analysis showed a significant main effect of task phase ($F(1,22) = 22.696$, $p < .001$) in which the mean reach angle at the end of the perturbation phase was significantly lower at baseline. Although the CB group learned slightly less, there was no main effect of group ($F(1,22) = .684$, $p = .417$) or group-by-phase interaction ($F(1,22) = .824$, $p = .374$). This result was corroborated by analysis of the total learning (Figure 2B), which was not significantly different between groups ($t(22) = -.908$, $p = .374$), and both groups showed total learning that was significantly greater than zero (CB: $t(11) = 2.645$, $p = .023$; Control: $t(11) = 4.142$, $p = .002$).

ANOVA testing differences in reach angle variability (i.e., standard deviation) between groups and across task phases showed no significant main effects of group ($F(1,22) = .452$, $p = .566$) or task phase ($F(1,22) = 3.814$, $p = .064$), or group-by-phase interaction ($F(1,22) = .460$, $p = .505$). Similarly, ANOVA comparing reach velocity across the groups and task phases showed no significant main effects or interaction (Figure 2C; group: $F(1,22) = .340$, $p = .566$; task phase: $F(1,22) = .516$; group-by-phase: $F(1,22) = .012$, $p = .914$). To quantify changes in reach velocity within subjects, we computed the difference in reach peak velocity between the baseline and end of perturbation phases (Figure 2D). The CB group showed larger between-subject variance than the Control group, with most CB participants reducing their speed over the course of the RL task. However, the difference between groups was not significant ($t(22) = -.110$, $p = .914$), and neither

group distribution was significantly different from zero (CB: $t(11) = -.465$, $p = .651$; Control: $t(11) = -.667$, $p = .519$).

Figure 3 shows the performance of both groups in the somatosensory state estimation task. ANOVA comparing the just noticeable difference (JND) values across groups and movement conditions showed a significant main effect of movement condition ($F(1,22) = 7.559$, $p = .012$; Figure 3A). The effect was driven by smaller JND values, indicating improved somatosensory acuity, in the active compared to the passive movement condition. There was no main effect of group ($F(1,22) = 2.262$, $p = .147$) or group-by-movement-condition interaction ($F(1,22) = .017$, $p = .898$), which indicated that both groups showed comparable performance. Indeed, this result was supported by the analysis of the active benefit (i.e., the difference in JND between the active and passive movement conditions) shown by each participant (Figure 3C). Means comparisons showed no significant difference in active benefit between groups ($t(22) = .129$, $p = .898$), and neither group distribution was significantly different from zero (CB: $t(11) = 2.101$, $p = .059$; Control: $t(11) = 1.799$, $p = .100$).

ANOVA comparing the somatosensory bias across groups and movement condition showed a main effect of movement condition, in which there was an overall negative bias in the active movement condition compared to approximately zero bias in the passive movement condition (Figure 3B; $F(1,22) = 5.236$, $p = .032$). There was no significant main effect of group ($F(1,22) = 2.094$, $p = .162$) or group-by-movement-condition interaction ($F(1,22) = 4.056$, $p = .056$), suggesting that both groups showed a comparable pattern of bias across the two movement conditions.

To quantify any differences in movement kinematics between groups in the active movement condition, we analyzed the peak velocity and the lateral trajectory variability (i.e., the standard deviation of position in the x-axis) of each reaching movement. Means comparisons showed that the CB group moved with greater velocity than the Control group on average ($t(22) = 2.222$, $p = .037$). However, the peak movement velocity did not show any relationship with the active benefit in either group (CB: $r = .189$, $p = .557$; CN: $r = .033$, $p = .918$). Despite the width of the force channel being somewhat large (5 cm), both groups showed a comparable degree of lateral movement ($t(22) = 1.383$, $p = .180$). Lateral movement variability also showed no relationship with active benefit (CB: $r = .032$, $p =.920$; CN: $r =.188$, $p = .559$).

The primary objective of this study was to assess the relationship between a measure of internal model integrity and reinforcement learning in the CB group. In particular, we aimed to assess the significance of this relationship in the presence of other potential predictors of learning. To this end, multiple linear regression was used to test if the following independent variables significantly predicted total learning in the RL task: the active benefit shown in the somatosensory state estimation task, the severity of CB participants' limb ataxia, the baseline variability shown in the RL task, the change in reach velocity in the RL task, and CB participants' age. The overall regression was statistically significant (R2 =.807, $F(5,6) = 5.014$, $p = .037$). Of all the independent variables included in the model, only the active benefit significantly predicted total learning in the RL task ($\beta = -.783$, $p = .007$; Figure 4). The complete results of the regression analysis are shown in Table 2.

DISCUSSION

In this study, we examined reinforcement motor learning (RL) and somatosensory state estimation in individuals with cerebellar degeneration (CB participants) as well as a group of neurotypical control participants (control participants) matched to CB participants for age and biological sex. The RL task was designed as a truncated version of the task from Therrien et al. (2016)[19] to isolate the early learning phase in which CB participants showed slower learning than controls. We found that the CB group learned slightly less than the control group in the present task, but the difference was not statistically significant. Overall, this replicates our prior results. The somatosensory state estimation task was modeled after the task used by Weeks et al. (2017b) [33], and was designed to permit calculation of the active benefit shown by each individual. Active benefit describes the improvement in somatosensory acuity that is typically seen when estimating active (i.e., voluntary) compared to passively induced movement of the body [27–29] and is thought to be subserved by cerebellum-dependent sensory prediction [31–33]. Here, CB and control groups showed comparable active benefit in the somatosensory state estimation task, which runs counter to previous work [31]. It is possible that the haptic wall, which served as a stop signal in the active movement condition, provided additional sensory input that aided performance. Prior literature has shown that individuals with cerebellar degeneration are able to use haptic feedback to compensate for their impairment [45]. Regardless, we found variance in the degree of active benefit across individuals, which aligns with the results of Weeks et al. (2017a) [32]. We found that active benefit was a significant predictor of RL in the CB group. Other predictors assessed – namely, ataxia severity, movement variability, the change in movement speed, and participant age – showed no significant

relationship with learning. Our results suggest that RL, while intact after cerebellar damage, is related to state estimation reliability in this population.

The regression analysis performed here constitutes an important test of the mathematical model proposed by Therrien et al. (2016) [19], which posited that in motor contexts, efficient RL depends on reliable state estimation in order to map reinforcement signals to the correct movement (i.e., solve the credit-assignment problem). The model was used to show that inefficient RL observed in individuals with cerebellar degeneration could be explained by impaired estimation of the movement performed on each trial. A similar idea has been proposed and studied by others [20,24]. Although the model in Therrien et al. (2016) was agnostic to the cause of impaired estimation, in the case of CB participants, the hypothesis was that cerebellar damage impaired the predictive component of state estimation. If this were true, one would expect RL to correlate with a behavioral measure of the predictive component of state estimation in this population. The present study demonstrates this correlation and provides evidence that impairments in predictive state estimation explain significantly more variance in RL across CB participants than other potential predictors. Overall, the results of this study support the model of Therrien et al. and speak to a mechanism by which the cerebellum plays an indirect role in RL by contributing to state estimation.

Yet, there remain additional factors that could explain total learning variance across individuals in our RL task. Studies of younger neurotypical individuals have shown that motor learning with binary reinforcement is supported in large part by the use of explicit cognitive strategies [35,36]. Although, it should be noted that RL has also been shown to support an implicit process in

samples with similar demographics [46–48], suggesting that learning with binary feedback likely reflects a mix of active ingredients. Cerebellar damage has been shown to disrupt the computation of explicit strategies in comparable motor learning tasks [37,38], but this disruption is difficult to disentangle from general cognitive declines associated with the advanced age of these study samples [39]. While we did not find a significant relationship between participant age and RL here, we did not quantitatively assess strategy use. It would be interesting for future work to directly test the relationship between explicit strategy computation and motor learning with binary reinforcement in individuals with cerebellar damage.

Variability in RL across CB participants has also been suggested to stem from a cerebellar role in the fundamental processing of reward information [49]. Indeed, there is evidence from studies of animal models that the cerebellum encodes information about appetitive and social rewards [50–53]. Studies in humans have also shown that cerebellar damage can disrupt learning from monetary rewards in non-motor tasks [54–56]. However, important differences exist between the cerebellar encoding of reward and the canonical reward encoding in midbrain dopaminergic neurons [50,57]. Furthermore, the relationship between elements of appetitive and monetary reward processing (e.g., reward utility, expected value computation, etc.) and learning with arbitrary binary feedback signals (e.g., an auditory tone) is unclear. Do individuals need to find binary feedback signals as fundamentally rewarding as monetary gain in order to effectively learn from them? Further studies are needed to answer this question.

The present results raise an interesting point regarding the potential utility of RL in designing novel motor rehabilitation therapies for CB participants. Our results suggest that the factor

underlying inefficient RL after cerebellar damage – impaired sensory prediction – is the same factor thought to underly impaired adaptation in this population. Impaired adaptation is posited to be an important contributor to the variable rehabilitation outcomes seen in individuals with cerebellar damage [58–60]. Thus, it may seem disadvantageous to study interventions leveraging RL. Despite inter-individual variance, Therrien et al. (2016) [19] showed that, on average, the same CB participants could learn with binary reinforcement but could not learn in adaptation conditions (evidenced by the respective presence and absence of after-effects). When taken with our current results, this suggests that there may be a gradient in dependence on sensory prediction across learning mechanisms [61,62]. The implication for CB participants is that there may be a critical period in which internal model function has degraded enough to preclude adaptation but is sufficiently spared to permit relatively efficient RL. A pilot study by Therrien et al. (2021) [63] showed that an RL intervention could improve the irregular trajectories that characterize cerebellar reaching ataxia. Although studies with larger sample sizes and additional statistical modeling are needed, it is possible that a test of active benefit could one day serve as a metric of which CB participants might benefit from an RL intervention.

CONCLUSION

This study provides evidence that effective reinforcement motor learning (RL) after cerebellar damage is related to the reliability of state estimation. We compared a sample of individuals with cerebellar degeneration (CB participants) and a sample of age- and sex-matched neurotypical controls (control participants) as they performed an RL task and a task that tested somatosensory estimates of limb state after active and passive movement. The somatosensory state estimation

task permitted assessment of active benefit – the degree of improvement in somatosensory acuity with active compared to passive movement – in each participant. Overall, the two groups performed similarly in both tasks. However, there was marked inter-individual variance, which aligned with prior literature. Within the cerebellar group, inter-individual variance in RL was best predicted by active benefit, suggesting a relationship between cerebellar-dependent sensory prediction and RL. Overall, the results of this study corroborate prior models proposing that the cerebellum supports RL efficiency by improving state estimation and helping to resolve the credit-assignment problem.

Funding

This work was supported by pilot project funding from the Moss Rehabilitation Research Institute (MRRI) Peer Review Committee and start-up funding from the MRRI awarded to AST.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contributions

AST conceived and designed the research, analyzed data, prepared figures, interpreted results, and wrote the manuscript. AST and ECS programmed the experimental tasks. CW and ECS performed the experiments and assisted with data analysis.

REFERENCES

1. Glickstein M, Strata P, Voogd J. Cerebellum: history. Neuroscience. 2009;162:549–59.

2. Wolpert DM, Miall RC, Kawato M. Internal models in the cerebellum. Trends Cogn Sci. 1998;2:338–47.

3. Wolpert DM, Ghahramani Z. Computational principles of movement neuroscience. Nat Neurosci. 2000;3:1212–7.

4. Krakauer JW, Pine ZM, Ghilardi MF, Ghez C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. J Neurosci. 2000;20:8916–24.

5. Shadmehr R, Mussa-Ivaldi FA. Adaptive representation of dynamics during learning of a motor task. J Neurosci. 1994;14:3208–24.

6. Bastian AJ, Martin TA, Keating JG, Thach WT. Cerebellar ataxia: abnormal control of interaction torques across multiple joints. J Neurophysiol. 1996;76:492–509.

7. Day BL, Thompson PD, Harding AE, Marsden CD. Influence of vision on upper limb reaching movements in patients with cerebellar ataxia. Brain. 1998;121 (Pt 2):357–72.

8. Bhanpuri NH, Okamura AM, Bastian AJ. Predicting and correcting ataxia using a model of cerebellar function. Brain. 2014;137:1931–44.

9. Weiner MJ, Hallett M, Funkenstein HH. Adaptation to lateral displacement of vision in patients with lesions of the central nervous system. Neurology. 1983;33:766–72.

10. Martin TA, Keating JG, Goodkin HP, Bastian AJ, Thach WT. Throwing while looking through prisms: I. Focal olivocerebellar lesions impair adaptation. Brain. 1996;119:1183–98.

11. Maschke M, Gomez CM, Ebner TJ, Konczak J. Hereditary cerebellar ataxia progressively impairs force adaptation during goal-directed arm movements. J Neurophysiol. 2004;91:230–8.

12. Morton SM, Bastian AJ. Prism adaptation during walking generalizes to reaching and requires the cerebellum. J Neurophysiol. 2004;92:2497–509.

13. Smith MA, Shadmehr R. Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. J Neurophysiol. 2005;93:2809–21.

14. Tseng Y, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ. Sensory prediction errors drive cerebellum-dependent adaptation of reaching. J Neurophysiol. 2007;98:54–62.

15. Izawa J, Criscimagna-Hemminger SE, Shadmehr R. Cerebellar contributions to reach adaptation and learning sensory consequences of action. J Neurosci. 2012;32:4230–9.

16. Schlerf JE, Xu J, Klemfuss NM, Griffiths TL, Ivry RB. Individuals with cerebellar degeneration show similar adaptation deficits with large and small visuomotor errors. J Neurophysiol. 2013;109:1164–73.

17. Haith AM, Krakauer JW. Model-based and model-free mechanisms of human motor learning. Adv Exp Med Biol. 2013;782:1–21.

18. Krakauer JW, Hadjiosif AM, Xu J, Wong AL, Haith AM. Motor learning. Compr Physiol. 2019;9:613–63.

19. Therrien AS, Wolpert DM, Bastian AJ. Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. Brain. 2016;139:101–14.

20. McDougle SD, Boggess MJ, Crossley MJ, Parvin D, Ivry RB, Taylor JA. Credit assignment in movement-dependent reinforcement learning. PNAS. 2016;113:6797–802.

21. Uehara S, Mawase F, Celnik P. Learning similar actions by reinforcement or sensory-prediction errors rely on distinct physiological mechanisms. Cereb Cortex. 2018;28:3478–90.

22. Niv Y. Reinforcement learning in the brain. J Math Psychol. 2009;53:139–54.

23. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. Cambridge, Massachusetts: The MIT Press; 1998.

24. Taylor JA, Ivry RB. Cerebellar and Prefrontal Cortex Contributions to Adaptation, Strategies, and Reinforcement Learning. In: Ramnani N, editor. Progress in Brain Research. Elsevier; 2014. p. 217–53.

25. Stengel RF. Optimal Control and Estimation. Dover Publications; 1994.

26. Wolpert DM, Ghahramani Z, Jordan MI. An internal model for sensorimotor integration. Science. 1995;269:1880–2.

27. Paillard J, Brouchon M. A proprioceptive contribution to the spatial encoding of position cues for ballistic movements. Brain Res. 1974;71:273–84.

28. Adamovich SV, Berkinblit MB, Fookson O, Poizner H. Pointing in 3D space to remembered targets. I. Kinesthetic versus visual target presentation. J Neurophysiol. 1998;79:2833–46.

29. Fuentes CT, Bastian AJ. Where is your arm? Variations in proprioception across space and tasks. J Neurophysiol. 2010;103:164–71.

30. Bhanpuri NH, Okamura AM, Bastian AJ. Active force perception depends on cerebellar function. J Neurophysiol. 2012;107:1612–20.

31. Bhanpuri NH, Okamura AM, Bastian AJ. Predictive modeling by the cerebellum improves proprioception. J Neurosci. 2013;33:14301–6.

32. Weeks HM, Therrien AS, Bastian AJ. Proprioceptive localization deficits in people with cerebellar damage. Cerebellum. 2017;16:427–37.

33. Weeks HM, Therrien AS, Bastian AJ. The cerebellum contributes to proprioception during motion. J Neurophysiol. 2017;118:693–702.

34. Zimmet AM, Cao D, Bastian AJ, Cowan NJ. Cerebellar patients have intact feedback control that can be leveraged to improve reaching. Elife. 2020;9:e53246.

35. Holland P, Codol O, Galea JM. Contribution of explicit processes to reinforcement-based motor learning. J Neurophysiol. 2018;119:2241–55.

36. Codol O, Holland PJ, Galea JM. The relationship between reinforcement and explicit control during visuomotor adaptation. Sci Rep. 2018;8:9121.

37. Butcher PA, Ivry RB, Kuo S-H, Rydz D, Krakauer JW, Taylor JA. The cerebellum does more than sensory prediction error-based learning in sensorimotor adaptation tasks. J Neurophysiol. 2017;118:1622–36.

38. Tsay JS, Schuck L, Ivry RB. Cerebellar degeneration impairs strategy discovery but not strategy recall. Cerebellum. 2022;

39. Wong AL, Marvel CL, Taylor JA, Krakauer JW. Can patients with cerebellar disease switch learning mechanisms to reduce their adaptation deficits? Brain. 2019;142:662–73.

40. Brand M, Schiebener J. Interactions of age and cognitive functions in predicting decision making under risky conditions over the life span. J Clin Exp Neuropsychol. 2013;35:9–23.

41. Eppinger B, Nystrom LE, Cohen JD. Reduced sensitivity to immediate reward during decision-making in older than younger adults. PLoS One. 2012;7:e36953.

42. Weller JA, King ML, Figner B, Denburg NL. Information use in risky decision making: Do age differences depend on affective context? Psychol Aging. 2019;34:1005–20.

43. Trouillas P, Takayanagi T, Hallett M, Currier RD, Subramony SH, Wessel K, et al. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. J Neurol Sci. 1997;145:205–11.

44. Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys. 2001;63:1293–313.

45. Gibo TL, Criscimagna-Hemminger SE, Okamura AM, Bastian AJ. Cerebellar motor learning: are environment dynamics more important than error size? J Neurophysiol. 2013;110:322–33.

46. Mawase F, Uehara S, Bastian AJ, Celnik P. Motor Learning Enhances Use-Dependent Plasticity. J Neurosci. 2017;37:2673–85.

47. van Mastrigt NM, Tsay JS, Wang T, Avraham G, Abram SJ, van der Kooij K, et al. Implicit reward-based motor learning. Exp Brain Res. 2023;

48. Forano M, Franklin DW. Reward actively engages both implicit and explicit components in dual force field adaptation. bioRxiv; 2023. p. 2023.08.09.552587.

49. Vassiliadis P, Derosiere G, Duque J. Beyond Motor Noise: Considering other causes of impaired reinforcement learning in cerebellar patients. eNeuro. 2019;6:ENEURO.0458-18.2019.

50. Hull C. Prediction signals in the cerebellum: Beyond supervised motor learning. eLife. 9:e54073.

51. Yoshida J, Oñate M, Khatami L, Vera J, Nadim F, Khodakhah K. Cerebellar contributions to the basal ganglia influence motor coordination, reward processing, and movement vigor. J Neurosci. 2022;42:8406–15.

52. 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.

53. Larry N, Yarkoni M, Lixenberg A, Joshua M. Cerebellar climbing fibers encode expected reward size. Elife. 2019;8:e46870.

54. Thoma P, Bellebaum C, Koch B, Schwarz M, Daum I. The Cerebellum Is Involved in Reward-based Reversal Learning. Cerebellum 2008;7:433-43.

55. Cardoso C de O, Branco LD, Cotrena C, Kristensen CH, Schneider Bakos DDG, Fonseca RP. The impact of frontal and cerebellar lesions on decision making: evidence from the Iowa Gambling Task. Front Neurosci. 2014;8:61.

56. Nicholas J, Amlang C, Lin C-YR, Montaser-Kouhsari L, Desai N, Pan M-K, et al. The role of the cerebellum in learning to predict reward: evidence from cerebellar ataxia. bioRxiv; 2022. p. 2022.11.04.515251.

57. Lixenberg A, Yarkoni M, Botschko Y, Joshua M. Encoding of eye movements explains reward-related activity in cerebellar simple spikes. J Neurophysiol. 2020;123:786–99.

58. Morton SM, Bastian AJ. Can rehabilitation help ataxia? Neurology. 2009;73:1818–9.

59. Ilg W, Bastian AJ, Boesch S, Burciu RG, Celnik P, Claaßen J, et al. Consensus paper: Management of degenerative cerebellar disorders. Cerebellum. 2014;13:248–68.

60. Ilg W, Timmann D. Gait ataxia—specific cerebellar influences and their rehabilitation. Movement Disorders. 2013;28:1566–75.

61. Izawa J, Shadmehr R. Learning from sensory and reward prediction errors during motor adaptation. PLoS Comput Biol. 2011;7:e1002012.

62. Cashaback JGA, McGregor HR, Mohatarem A, Gribble PL. Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. PLoS Comput Biol. 2017;13:e1005623.

63. Therrien AS, Statton MA, Bastian AJ. Reinforcement signaling can be used to reduce elements of cerebellar reaching ataxia. Cerebellum. 2021;20:62–73.

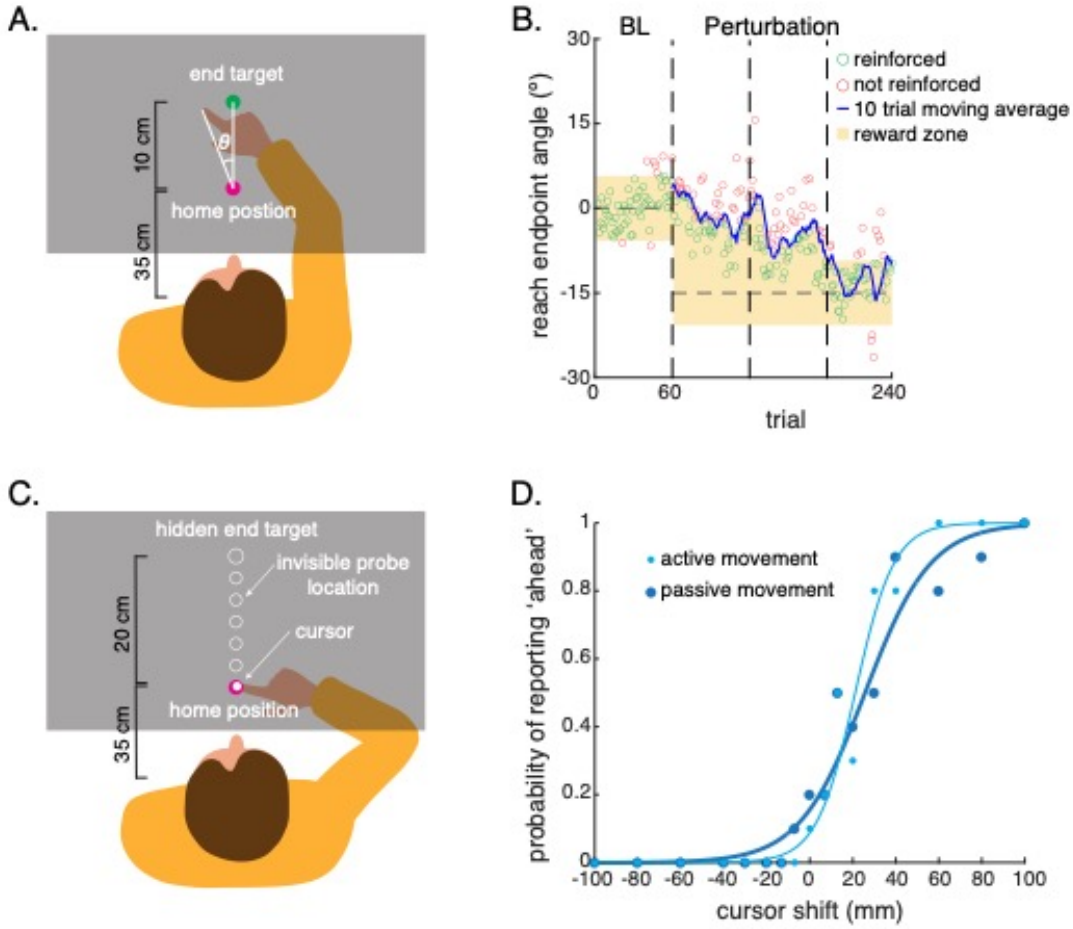

Figure 1. A schematic of the reinforcement motor learning (RL) and somatosensory state estimation tasks. Both tasks were performed with the dominant arm. **A.** The RL task had participants make 10 cm reaching movements from a home position to a single end target. The task drove participants to learn to rotate the endpoint angle of their reaches. The reach angle was defined as the angle between one vector connecting the home position to the end target and another connecting the home position to the index fingertip position when it crossed a radial distance of 10 cm from the home position (i.e., the reach endpoint). Throughout the task, the arm moved below a screen and was hidden from view. Participants received binary feedback about the outcome of their reaches. Reaches that successfully crossed a reward zone were reinforced with the target turning green and an auditory tone. Reaches that did not cross the reward zone resulted in the target turning red and no auditory tone. **B.** Data from an example individual with cerebellar degeneration showing the trial and feedback structure of the RL task. After a baseline phase in which the reward zone corresponded to the visually displayed target, closed-loop reinforcement feedback was used to drive participants to rotate their reach angle. A running average of the reach angle from the previous 10 reaches was computed. Reaches with an angle that fell between the running average and the rotated target were reinforced. Reach angles that fell outside this range were not reinforced. **C.** In the somatosensory state estimation task, participants actively moved or were passively moved along a 20 cm reach trajectory. A visual cursor showed the index fingertip position for a brief period at the beginning of each trial before

being extinguished. On each trial, the hand was stopped at 1 of 5 possible probe locations. A visual cursor reappeared for a 600 ms display period, but its position was shifted relative to the true position of the index finger. After 600 ms, the cursor disappeared, and participants completed the reach. Participants then reported whether they perceived the cursor as 'ahead' of or 'behind' the index fingertip during the display period in a 2-alternative forced choice paradigm. **D.** Data from the same example participant shown in panel B. The probability of reporting 'ahead' was determined for each of the 17 cursor shift values tested. The data from each movement condition were then fit with separate psychometric functions, from which measures of somatosensory acuity and bias were derived.

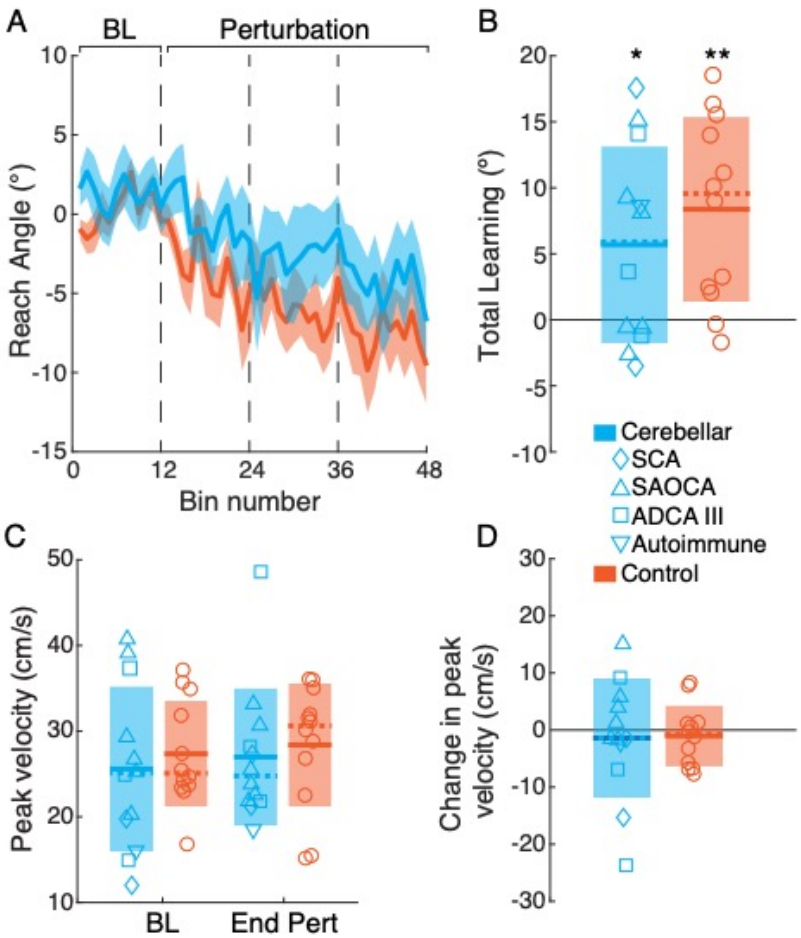

Figure 2. Behavior in the reinforcement motor learning task. **A.** The time series of reach angles shown by the cerebellar and control groups. Solid lines represent the group mean, and the shaded regions depict the standard deviation. Each bin comprises the mean of 5 trials. **B.** The total learning (i.e., the difference in mean reach angle between the end of the baseline and perturbation phases) shown by each group. **C.** The mean reach peak velocity shown by each group at the end of the baseline (BL) and perturbation (End Pert) phases. **D.** The within-subject change in reach peak velocity between the end of the baseline and perturbation phases. **B-D.** Markers represent individual participants. Solid lines represent the group mean. Dashed lines represent the group median. The shaded regions depict the standard deviation. SCA: Spinocerebellar ataxia. SAOCA: Sporadic adult-onset ataxia. ADCA III: Autosomal dominant cerebellar ataxia type 3. Autoimmune: Autoimmune mediated cerebellar degeneration. * p <.05 ** p <.01.

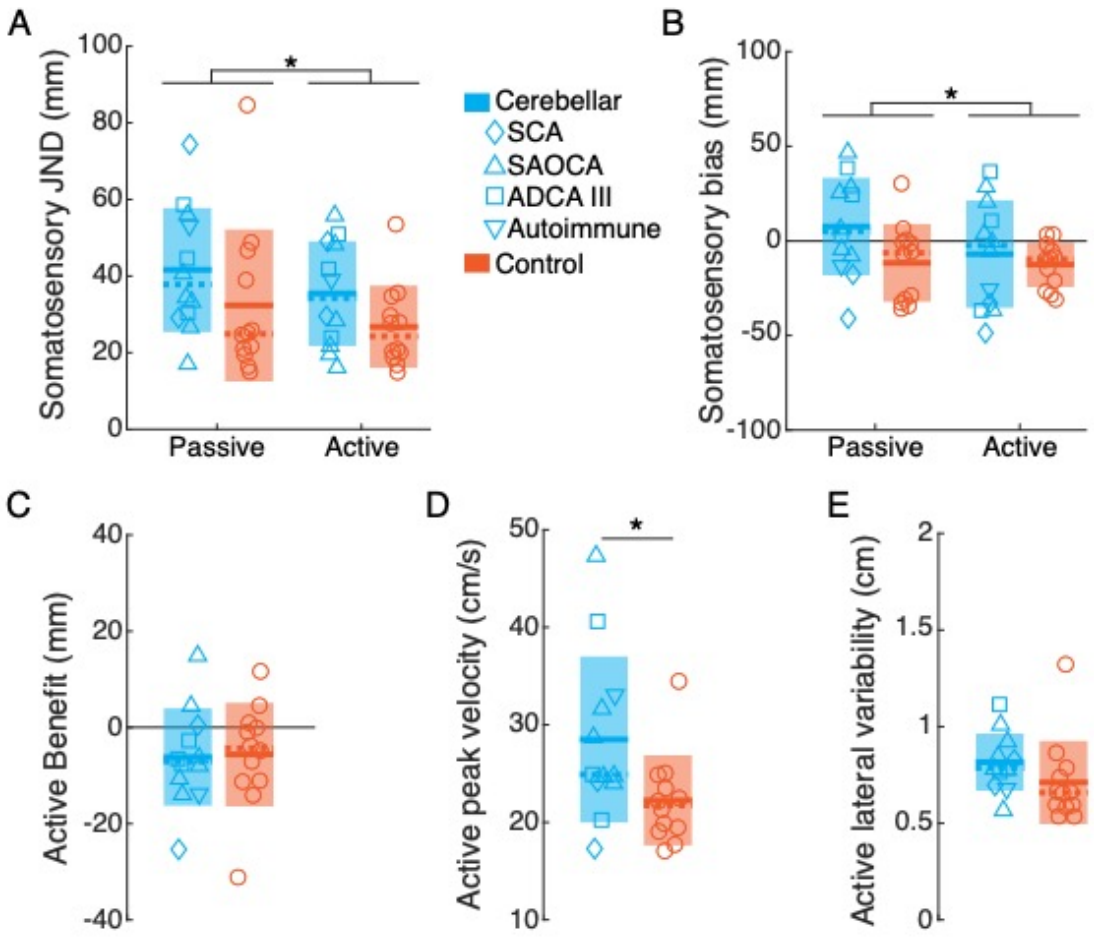

Figure 3. Behavior in the somatosensory state estimation task. **A-B.** The just noticeable difference (JND, **A**) and the bias (**B**) shown by the cerebellar and control groups in the passive and active movement conditions. **C.** The active benefit (i.e., the within-subject difference in JND between the active and passive movement conditions) shown by each group. **D.** The mean peak velocity of reaching movements in the active movement condition shown by each group. **E.** The standard deviation of position along the x-axis of reaching movements in the active movement condition shown by each group. **A-E.** Markers represent individual participants. Solid lines represent the group mean. Dashed lines represent the group median. The shaded regions depict the standard deviation. SCA: Spinocerebellar ataxia. SAOCA: Sporadic adult-onset ataxia. ADCA III: Autosomal dominant cerebellar ataxia type 3. Autoimmune: Autoimmune mediated cerebellar degeneration. * p <.05

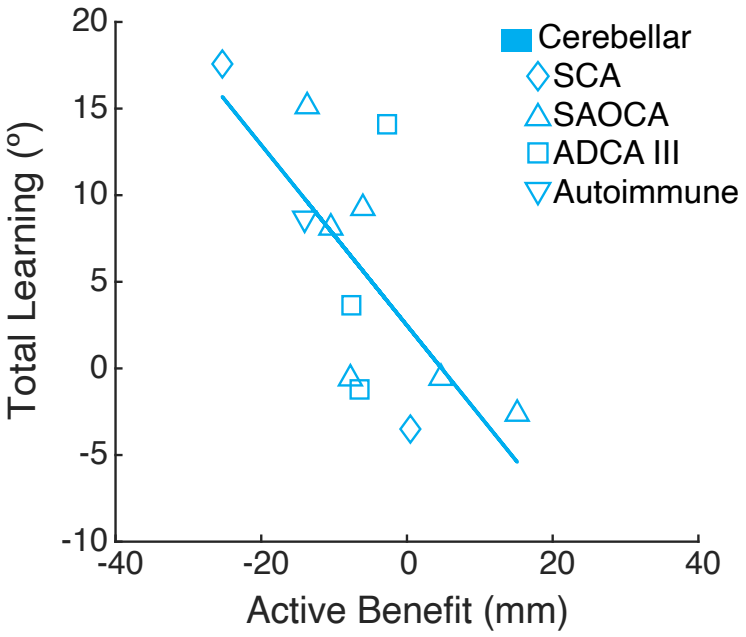

Figure 4. The relationship between the magnitude of active benefit in the somatosensory state estimation task and the total learning in the reinforcement learning task in the cerebellar group. The regression line depicts the bivariate regression for illustrative purposes. See Table 2 for the complete results of the regression analysis performed. SCA: Spinocerebellar ataxia. SAOCA: Sporadic adult-onset ataxia. ADCA III: Autosomal dominant cerebellar ataxia type 3. Autoimmune: Autoimmune mediated cerebellar degeneration.

Table 1. Participant Demographics.

Participant	Age (years)	Sex (M/F)	Handedness (R/L)	Diagnosis	ICARS	
					Total (/100)	Limb Kinetic (/52)
CB01	46	F	R	ADCAIII	41	17
CB02	23	M	R	SAOCA	46	23
CB03	75	M	R	SCA14	50	23
CB04	58	F	L	SCA8	23	8
CB05	59	M	R	SAOCA	52	28
CB06	58	F	R	ADCAIII	24	7
CB07	69	F	R	SAOCA	31	15
CB08	37	M	R	ADCAIII	25	11
CB09	55	M	R	SAOCA	42	23
CB10	73	M	R	Autoimmune	45	16
CB11	42	M	R	SAOCA	31	13
CB12	66	F	R	SAOCA	55	20
CB Group	**Mean: 55.1**	**M: 7**	**R: 11**		**Mean: 37.8**	**Mean: 16.6**
	SD: 15.6	**F: 5**	**L: 1**		**SD: 10.9**	**SD: 6.3**
Control Group	**Mean: 55.8**	**M: 7**	**R: 11**			
	SD: 15.8	**F: 5**	**L: 1**			

CB: Participant with cerebellar degeneration. Control: Neurotypical. M: Male; F: Female. R: Self-reported right-hand dominant. L: Self-reported left hand dominant. ADCAIII: Autosomal dominant cerebellar ataxia type 3. SAOCA: Sporadic adult-onset cerebellar ataxia. SCA: Spinocerebellar ataxia types 8 and 14. ICARS: International Cooperative Ataxia Rating Scale. The ICARS Total reflects the total score out of a possible 100. The ICARS Limb Kinetic reflects the limb coordination sub-score out of a possible 52. Higher scores on the ICARS indicate more severe ataxia.

Table 2. Results of multiple linear regression analysis to assess predictors of reinforcement learning in participants with cerebellar degeneration.

Predictor	β	Std. Error	t	p
ICARS limb kinetic score	.357	.216	1.650	.150
Baseline reach variability in the RL Task	-1.290	.999	-1.291	.244
Change in reach peak velocity in the RL task	-.228	.131	-1.736	.133
Participant age	-.251	.127	-1.969	.097
Active benefit	-.783	.197	-3.968	.007 **
Constant	15.056			
R^2	.807			
F	5.014			
SEE	4.435			
n	12			

** $p < .01$. ICARS: International Cooperative Ataxia Rating Scale. RL: Reinforcement learning.