The mid-lateral cerebellum is necessary for reinforcement learning

Naveen Sendhilnathan^{1-3#} & Michael E. Goldberg²⁻⁵

¹Doctoral program in Neurobiology and Behavior,

²Dept. of Neuroscience, Mahoney Center for Brain and Behavior Research, Zuckerman Mind, Brain, and Behavior Institute, Columbia University, New York, NY;

³New York State Psychiatric Institute, New York, NY;

⁴Kavli Institute for Brain Science, Columbia University, New York, NY;

⁵Dept. of Neurology, Psychiatry, and Ophthalmology, Columbia University College of Physicians and Surgeons, New York, NY;

#Correspondence:

Naveen Sendhilnathan,
3227 Broadway, L6-033
New York, NY 10027

ns3046@columbia.edu

Summary: The cerebellum has long been considered crucial for supervised motor learning and its optimization¹⁻³. However, new evidence has also implicated the cerebellum in reward based learning⁴⁻⁸, executive function⁹⁻¹², and frontal-like clinical deficits¹³. We recently showed that the simple spikes of Purkinje cells (P-cells) in the mid-lateral cerebellar hemisphere (Crus I and II) encode a reinforcement error signal when monkeys learn to associate arbitrary symbols with hand movements⁴. However, it is unclear if the cerebellum is necessary for any process beyond motor learning. To investigate if the mid-lateral cerebellum is actually necessary for learning visuomotor associations, we reversibly inactivated the mid-lateral cerebellum of two primates with muscimol while they learned to associate arbitrary symbols with hand movements. Here we show that cerebellar inactivation impaired the monkey's ability to learn new associations, although it had no effect on the monkeys' performance on a task with overtrained symbols. A computational model corroborates our results. Cerebellar inactivation increased the reaction time, but there were no deficits in any motor kinematics such as the hand movement, licking or eye movement. There was no loss of function when we inactivated a more anterior region of the cerebellum that is implicated in motor control. We suggest that the mid-lateral cerebellum, which provides a reinforcement learning error signal⁴, is necessary for visuomotor association learning. Our results have implications for the involvement of cerebellum in cognitive control, and add critical constraints to brain models of non-motor learning^{14,15}.

We trained two monkeys to associate arbitrary visual symbols with left and right hand movements to earn an immediate liquid reward. On each recording session, the monkeys first performed 30-50 trials of the task with familiar, overtrained symbols (**Fig 1a**) at a success rate close to 100%. We created a library of 16 pairs of fractal symbols with varying levels of similarities between the symbols. After the monkeys performed a few trials on the overtrained task (OT), we presented them with one of the novel symbol pairs from the library (**Fig 1a**) and tracked their learning. To calculate the monkeys' acquisition-difficulty level for different symbol pairs, we recorded the monkeys' performance for each pair (Monkey B- 16 pairs, Monkey S- 9 pairs) at least three times, spread over several weeks, presented in random order (**Fig 1b**). To prevent the monkeys from relying on previous experience with the same symbol pair to relearn their associations, we either presented a given symbol pair again only several days after its prior

presentation, with intervening presentations of other symbol pairs or random pairs not from the library (**Fig 1b**; symbol pair highlighted in yellow), or we reversed the symbol-hand association the next time we presented the same pair (**Fig 1b**; symbol pair highlighted in red). We collected three repetitions of behavioral performance for each symbol pair from both monkeys (Monkey B- 16 pairs, Monkey S- 9 pairs). We defined the acquisition-difficulty level for a given symbol pair, for a specific monkey, as the trial at which the monkeys' performance was at least 75% accurate (see methods). The acquisition-difficulty rates were similar for each presentation of a given symbol pair (**Fig 1c**) indicating that the monkeys inevitably forgot their prior experience with a given symbol pair during each additional re-learning.

Although P-cells in the mid-lateral cerebellum, encode a reinforcement learning error signal⁴, it is not clear if this activity is only correlative or causal to behavior. To test if mid-lateral cerebellum is necessary for the monkeys to perform the task, we infused 4-10 μ l of a 10 mg/ml solution of muscimol, a GABA_A agonist that hyperpolarizes cell bodies without affecting fibers of passage, near crus I and II of the mid-lateral cerebellum of two monkeys in the same location where we had recorded task-dependent P-cells (see methods, **Fig 2a**; see **Fig S1** for infusion locations). We infused saline alone on the days before and after every muscimol injection as a control.

The monkeys typically learned the novel association in 50-70 trials during the control sessions (**Fig 2c, 2f**). Mid-lateral cerebellar inactivation prevented them from learning the correct association for the same pairs of symbols from the library that we used for the saline condition. In the beginning of learning, the behavioral performances for control and inactivation conditions were comparable, with low mean squared error (Monkey B: $P = 0.1117$; t-test; **Fig 2d**; Monkey S: $P = 0.2196$; paired t-test; **Fig 2g**). However, after 75 trials of learning, when the monkeys performed more than 75% correct with low mean squared error during control conditions, their performance was close to chance level (Monkey B: $P = 2.6410 \times 10^{-5}$; t-test; **Fig 2d**; Monkey S: $P = 4.1135 \times 10^{-5}$; ranksum test; **Fig 2g**) with significantly higher mean squared error during the inactivation condition (Monkey B: $P = 3.3216 \times 10^{-4}$; ranksum test; **Fig 2d**; Monkey S: $P = 0.0016$; ranksum test; **Fig 2g**). Cerebellar inactivation affected the bias more than the variance during learning (**Fig S2**). The learning performance during cerebellar inactivation was independent of the rate of learning of a symbol pair in the control condition for both monkeys. When the cerebellum was inactivated, the monkeys could not even learn the association that had the least acquisition difficulty (**Fig 2e**,

2h). Repetition of symbols after error trials during learning significantly improved the performance (rate of learning) under control-saline condition ($P = 0.0029$; paired t-test) but not during cerebellar inactivation ($P = 0.1117$; paired t-test) for the same number of trials, for the same symbol pairs (**Fig S3**). We verified that the performance modulations across sessions, within a day, did not confound our observation (**Fig S4**).

Mid-lateral cerebellar inactivation did not alter the monkeys' behavioral performance on the overtrained task: The monkeys continued to perform close to 100% accuracy when the cerebellum was inactivated, similar to their performance during the saline-control condition (Monkey B: $P = 0.2110$; Mann-Whitney U test; **Fig 2c**; Monkey S: $P = 0.9132$; Mann-Whitney U test; **Fig 2f**). The inactivation did not affect the mean squared error (Monkey B: $P = 0.2518$; Mann-Whitney U test; **Fig 2d**; Monkey S: $P = 0.2787$; Mann-Whitney U test; **Fig 2g**) in the performance, the decomposed bias squared (**Fig S2**) or the variance (**Fig S2**) during the overtrained task.

Although inactivation of mid-lateral cerebellum prevented the monkeys from learning the visuomotor association for a given hand movement, it did not affect the gross motor kinematics of hand movement: In both control and inactivation conditions, the monkeys performed the task with well-stereotyped hand movements similar to those in the OT (**Fig 3a**; Monkey B: $P = 0.7342$; Mann-Whitney U test; Monkey S: $P = 0.5286$; paired t-test) and during learning (**Fig 3b**; Monkey B: $P = 0.8703$; paired t-test; Monkey S: $P = 0.3195$; paired t-test). The licking behavior was also not different between the control and inactivation conditions for OT (**Fig 3c**; Monkey B: $P = 0.7631$; paired t-test; Monkey S: $P = 0.9965$; paired t-test) or during learning (**Fig 3d**; Monkey B: $P = 0.4413$; paired t-test; Monkey S: $P = 0.4217$; paired t-test). Although monkeys generally tended to explore their environment more during learning than during OT, there were no systematic differences in their eye movements between the control and inactivation conditions during OT (**Fig 3e**; Monkey B: $P = 0.2280$; Mann-Whitney U test; Monkey S: $P = 0.7311$; paired t-test) or during learning (**Fig 3f**; Monkey B: $P = 0.1127$; Mann-Whitney U test; Monkey S: $P = 0.6665$; Mann-Whitney U test). Additionally, cerebellar inactivation did not have a differential effect on motor kinematics for correct and wrong trials (**Fig S6**). Furthermore, with cerebellar inactivation, the monkeys had no deficit performing the OT task with very different hand movements. We switched the-manipulanda that the monkeys used to reporting their choices, from bars to dowels. Although monkeys had to make an entirely different hand movement to report the same choices (**Fig S7**) cerebellar inactivation had no effect on the kinematics of the changed movement (**Fig S7**). The

only motor effect of the inactivation was to increase monkeys' manual reaction time for both monkeys during learning ($P = 0.0013$; paired t-test; **Fig 3g, h**). Given that there were no changes in any other aspect of the movements with which the monkeys signaled their decision, we inferred that inactivating the midlateral cerebellum interfered with the monkeys' ability make a decision about the association than their ability to report it.

Finally, to test if the effect of muscimol inactivation were limited to the mid-lateral cerebellum, we made muscimol injections into more anterior parts of the cerebellum (lobule V, an area thought to be involved in motor control¹⁶) in both monkeys (**Fig S9a-b**). We again recorded the monkeys' performance for the some of the same symbol pairs from the library, used for prior experiments (Monkey B- 5 pairs, Monkey S- 4 pairs). Inactivation of the anterior cerebellum had no effect on the monkey's ability to learn new visuomotor associations or to perform the overtrained association (**Fig S9c-g**). However, it caused transient hind limb ataxia and nystagmus in one monkey and slightly reduced the monkeys' reaction time (**Fig S9e-g**).

The mid-lateral cerebellum, but not more anterior regions, lies within a closed loop network for non-motor learning, due to its anatomical connections with the basal ganglia and the prefrontal cortex^{14,17-19}. Neurons in the prefrontal cortex²⁰⁻²² the globus pallidus internal segment (GPi)²³, caudate nucleus report the progress of visuomotor association²¹. Lesions in prefrontal cortex of the monkey eliminated the ability of monkeys to learn a new visuomotor association and impaired their ability to retain associations learned preoperatively²⁴. Combined lesions of basal ganglia and premotor cortex prevented monkeys from retaining associations learned preoperatively, but did not affect their ability to learn new associations²⁵.

We recently showed that the simple spikes of P-cells in the mid-lateral cerebellum encode reinforcement error signals during visuomotor association learning⁴. Two types of P-cells report whether the monkey's decision on the most recent trial was correct (correct-preferring neurons, cP-cells) or wrong (wrong-preferring neurons, wP-cells), and although each neuron reports this only during a brief epoch (the 'delta epoch'), the population as a whole maintains the information from shortly after the previous outcome to shortly after the next. The magnitude of this error signal is greatest at the beginning of learning, and approaches zero as the monkey learns the task, which is what is expected of a reinforcement learning error signal. The P-cells have no information about what the monkey will do in the upcoming trial or about past trials other than the most recent. When the monkeys perform the overtrained task, the neurons do not signal the occasional wrong decision.

In keeping with this, reversible inactivation of this area eliminates the ability of the monkey to learn new associations, but does not affect their ability to remember the overtrained association. Inactivation increases the reaction time of the hand movement, but the absence of any other kinematic effect suggests that this increase in reaction time arises from a difficulty the monkeys have in making a decision, rather than a motor deficit per se.

A similar dissociation between short-term and chronic learning has been demonstrated in the motor system. Cats exposed to 60 minutes of exposure to .25x miniaturizing spectacles reduce the gain of their vestibuloocular reflex(VOR) to 0.75 - 0.8 from 1²⁶. Injection of CQNX, an antagonist of the glutamate AMPA and kainite receptors into the flocculus, the vestibulocerebellum, eliminated this gain reduction. If the cats wore the spectacles for 3 days, and had three hour-long training sessions, their gain fell to 0.65, but CQNX injection increased it only slightly, to 0.75. These data implied that the long-term gain reduction was not exclusively dependent upon the cerebellum, but migrated elsewhere as learning proceeded. In the case of the VOR, this extracerebellar locus is the medial vestibular nucleus²⁷. In the case of reinforcement learning, the synaptic changes associated with learning that begin in the cerebellum might migrate to the prefrontal cortex, the basal ganglia, or both¹⁸. This still needs to be investigated.

References:

- 1 Raymond, J. L. & Medina, J. F. Computational Principles of Supervised Learning in the Cerebellum. *Annual review of neuroscience* **41**, 233-253 (2018).
- 2 Thach, S. G. L. W. T. in *Principles of Neural Science* Vol. 5 (ed Eric R. Kandel; James H. Schwartz; Thomas M. Jessell; Steven A. Siegelbaum; A. J. Hudspeth) Ch. 42, 960-981 (McGraw-Hill Education / Medical, 2012).
- 3 Marr, D. A theory of cerebellar cortex. *The Journal of Physiology* **202**, 437-470, doi:10.1113/jphysiol.1969.sp008820 (1969).
- 4 Sendhilnathan, N., Ipata, A. E. & Goldberg, M. E. Neural correlates of reinforcement learning in midlateral cerebellum. *Neuron* **106** (2020).
- 5 Heffley, W. & Hull, C. Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. *eLife* **8**, doi:10.7554/eLife.46764 (2019).
- 6 Heffley, W. *et al.* Coordinated cerebellar climbing fiber activity signals learned sensorimotor predictions. *Nature neuroscience* **21**, 1431-1441, doi:10.1038/s41593-018-0228-8 (2018).
- 7 Kostadinov, D., Beau, M., Pozo, M. & Häusser, M. Predictive and reactive reward signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. *Nature Neuroscience* **22**, 950-962, doi:10.1038/s41593-019-0381-8 (2019).
- 8 Sendhilnathan, N., Ipata, A. E. & Goldberg, M. E. Complex spikes encode reward expectation signals during visuomotor association learning. *BiorXiv* (2019).
- 9 Koziol, L. F., Budding, D. E. & Chidekel, D. From movement to thought: executive function, embodied cognition, and the cerebellum. *The Cerebellum* **11**, 505-525 (2012).
- 10 Koziol, L. F. *et al.* Consensus paper: the cerebellum's role in movement and cognition. *The Cerebellum* **13**, 151-177 (2014).
- 11 Wagner, M. J. & Luo, L. Neocortex–Cerebellum Circuits for Cognitive Processing. *Trends in Neurosciences* (2019).
- 12 Buckner, R. L. The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron* **80**, 807-815 (2013).
- 13 Schmahmann, J. D., MacMore, J. & Vangel, M. Cerebellar stroke without motor deficit: clinical evidence for motor and non-motor domains within the human cerebellum. *Neuroscience* **162**, 852-861 (2009).
- 14 Bostan, A. C. & Strick, P. L. The basal ganglia and the cerebellum: nodes in an integrated network. *Nature Reviews Neuroscience* **19**, 338-350 (2018).
- 15 Ito, M. Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience* **9**, 304-313 (2008).
- 16 Snider, R. S. & Eldred, E. Cerebro-cerebellar relationships in the monkey. *Journal of Neurophysiology* **15**, 27-40 (1952).
- 17 Kelly, R. M. & Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **23**, 8432-8444 (2003).
- 18 Strick, P. L., Dum, R. P. & Fiez, J. A. Cerebellum and nonmotor function. *Annual review of neuroscience* **32**, 413-434, doi:10.1146/annurev.neuro.31.060407.125606 (2009).
- 19 Caligiore, D. *et al.* Consensus Paper: Towards a Systems-Level View of Cerebellar Function: the Interplay Between Cerebellum, Basal Ganglia, and Cortex. *Cerebellum (London, England)* **16**, 203-229, doi:10.1007/s12311-016-0763-3 (2017).

- 186 20 Chen, L. L. & Wise, S. P. Neuronal activity in the supplementary eye field during
187 acquisition of conditional oculomotor associations. *Journal of Neurophysiology* (1995).
- 188 21 Pasupathy, A. & Miller, E. K. Different time courses of learning-related activity in the
189 prefrontal cortex and striatum. *Nature* **433**, 873-876, doi:10.1038/nature03287 (2005).
- 190 22 Histed, M. H., Pasupathy, A. & Miller, E. K. Learning substrates in the primate prefrontal
191 cortex and striatum: sustained activity related to successful actions. *Neuron* **63**, 244-253,
192 doi:10.1016/j.neuron.2009.06.019 (2009).
- 193 23 Inase, M., Li, B. M., Takashima, I. & Iijima, T. Pallidal activity is involved in visuomotor
194 association learning in monkeys. *Eur J Neurosci* **14**, 897-901, doi:10.1046/j.0953-
195 816x.2001.01701.x (2001).
- 196 24 Bussey, T. J., Wise, S. P. & Murray, E. A. The role of ventral and orbital prefrontal
197 cortex in conditional visuomotor learning and strategy use in rhesus monkeys (*Macaca*
198 *mulatta*). *Behavioral neuroscience* **115**, 971 (2001).
- 199 25 Nixon, P. D. & Passingham, R. The cerebellum and cognition: cerebellar lesions impair
200 sequence learning but not conditional visuomotor learning in monkeys.
201 *Neuropsychologia* **38**, 1054-1072 (2000).
- 202 26 Kassardjian, C. D. *et al.* The site of a motor memory shifts with consolidation. *Journal of*
203 *Neuroscience* **25**, 7979-7985 (2005).
- 204 27 Lisberger, S. G. Cerebellar LTD: a molecular mechanism of behavioral learning? *Cell* **92**,
205 701-704 (1998).
- 206 28 Hays, A. V., Richmond, B. J. & Optican, L. M. A UNIX-based multiple process system
207 for real-time data acquisition and control. *WESCON Conf. Proc.* **2**, 1-10 (1982).
- 208 29 Liu, Y., Yttri, E. A. & Snyder, L. H. Intention and attention: different functional roles for
209 LIPd and LIPv. *Nature neuroscience* **13**, 495 (2010).
- 210 30 Arikan, R. *et al.* A method to measure the effective spread of focally injected muscimol
211 into the central nervous system with electrophysiology and light microscopy. *Journal of*
212 *neuroscience methods* **118**, 51-57 (2002).
- 213 31 Tinevez, J.-Y. Y. *et al.* TrackMate: An open and extensible platform for single-particle
214 tracking. *Methods (San Diego, Calif.)* **115**, 80-90, doi:10.1016/j.ymeth.2016.09.016
215 (2017).
- 216 32 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature*
217 *methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

Acknowledgments:

We thank John Caban for making injectrodes and all other creative machining, Glen Duncan for highly creative and wonderful electronic assistance, Dr. Girma Asfaw, Dr. Moshe Shalev for animal care, Whitney Thomas and Holly Cline for facilitating everything. This work was supported by the Keck, Zegar Family, and Dana Foundations and the National Eye Institute (R24 EY-015634, R21 EY-017938, R21 EY-020631, R01 EY-017039, and P30 EY-019007 to M. E. Goldberg, PI).

Raw data and computer codes are available upon request.

Figure 1: Visuomotor association learning task paradigm

- Two-alternative forced-choice discrimination task (top) and parameters (bottom): The monkeys pressed each bar with a hand, and a white square appeared that served as the cue for the start of the trial. Then one of the two visual symbols appeared briefly. The monkey lifted the hand associated with the presented symbol to earn a reward immediate. Left: a trial from overtrained condition and right: a trial from novel learning condition.
- Trial session organization. Sessions marked with yellow rectangle indicate repeated presentation of the same symbol pairs. Sessions marked with red rectangle indicate repeated presentation of the same symbol pairs but with reversed visuomotor association.
- Acquisition-difficulty for each symbol pair (colored lines) per repetition, for both monkeys. Acquisition-difficulty for the second learning of the same symbol pair was not different from the first learning for either monkeys (monkey B: $P = 0.8608$; monkey S: $P = 0.5489$ paired t-test). Similarly, Acquisition-difficulty for the third learning was not different from the second learning for either monkeys (monkey B: $P = 0.2640$; monkey S: $P = 0.5228$ paired t-test).

Figure 2: Cerebellar inactivation caused an impairment in visuomotor association learning

- Schematic illustration of the mid-lateral cerebellum inactivation and simultaneous recording.
- Schematic illustration of three possible ways in which cerebellar inactivation (yellow) can affect the learning curve compared to control conditions (gray).
- Behavioral performance of Monkey B during OT and learning trials for control-saline (gray) and cerebellar inactivation (yellow) conditions.
- Quantitation of mean squared error ($MSE = \text{bias}^2 + \text{variance}$) from **Fig 2c**; OT: $P = 0.02518$, ranksum test; L_{beg} : $P = 0.1117$, paired t-test; L_{end} : $P = 2.6410 \times 10^{-5}$, paired t-test. For the bias^2 and variance, see **Fig S2**.
- Left: heat plot of learning curves (each row) for the behavioral performances of 16 pairs of symbols, arranged in the increasing order of acquisition difficulty during the saline-control condition. Right: heat plot of learning curves (each row) for the behavioral performances of the same 16 pairs of symbols (in the same order of left panel) during cerebellar inactivation.
- Same as **Fig 2c**, but for monkey S.
- Quantitation mean squared error (MSE) from **Fig 2g**; OT: $P = 0.2787$, ranksum test; L_{beg} : $P = 0.2196$, paired t-test; L_{end} : $P = 0.0016$, ranksum test. For the bias^2 and variance, see **Fig S2**.
- Same as **Fig 2e**; but for Monkey S.

Figure 3: Cerebellar inactivation did not cause any deficit in motor kinematics

- Left: Average horizontal (H) and vertical (V) hand positions across sessions for monkey B, during OT for control-saline (gray) and muscimol-inactivation (yellow) conditions. See **Fig S5** for Monkey S. Right: Quantitation of the amplitude of hand movement for monkey B (left; circle markers; $P = 0.7342$, ranksum test) and monkey S (right, square markers; $P = 0.5286$; paired t-test).
- Same as **Fig 3a**; but during learning condition. Right panel, monkey B (left; circle markers; $P = 0.8703$, paired t-test) and monkey S (right, square markers; $P = 0.3195$; paired t-test).
- Left: Average licking activity across sessions and monkeys during OT for control-saline (gray) and muscimol-inactivation (yellow) conditions. Right: Quantitation for monkey B (left; circle markers; $P = 0.7631$, paired t-test) and monkey S (right, square markers; $P = 0.9965$; paired t-test).
- Same as **Fig 3c**; but during learning condition. Right panel, monkey B (left; circle markers; $P = 0.4413$, paired t-test) and monkey S (right, square markers; $P = 0.4217$; paired t-test).
- Left: Average eye movement exploration activity on the full screen, across sessions and monkeys during OT for control-saline (gray) and muscimol-inactivation (yellow) conditions. Heat map is shown in the inset. Right: Quantitation of radius of visual exploration (r_{exp} ; see methods) for monkey B (left; circle markers; $P = 0.2280$, ranksum test) and monkey S (right, square markers; $P = 0.7311$; paired t-test).
- Same as **Fig 3e**; but during learning condition. Right panel, monkey B (left; circle markers; $P = 0.1127$, ranksum test) and monkey S (right, square markers; $P = 0.6665$; ranksum test).
- Average RT profile during OT and learning across sessions and monkeys for control-saline (gray) and inactivation (yellow) conditions.
- Quantitation from **Fig 3g** for two monkeys separately. Left: Monkey B, OT: $P = 0.4774$; paired t-test; L_{beg} : $P = 0.1011$; paired t-test; L_{end} : $P = 3.1909 \times 10^{-4}$ paired t-test. Right: Monkey S, OT: $P = 0.3028$; paired t-test; L_{beg} : $P = 0.2826$; paired t-test; L_{end} : $P = 0.0021$ paired t-test.

Figure S1: Recording and infusion locations

T1 MRI of the cerebellum infusion locations (black markers) in the coronal view (left) and sagittal view (right) for monkeys B and S. Yellow shading region surrounding the black markers represent the extent of Muscimol diffusion-spread per injection. The region marked by the white broken line is the projected region to contain P-cells that carry reinforcement error signal based on prior recordings.

Figure S2: Cerebellar inactivation increased bias and variance in performance during learning but during OT condition

- Changes in bias² (left) and changes in variance (right) for Monkey B from control-saline (S, gray) to Muscimol (M, yellow) condition during OT, beginning of learning (L_{beg}) and end of learning (L_{end}).
- Same as **Fig S2a**, but for Monkey S.

Figure S3: Repetition of symbols after error trials during learning significantly improved the performance during control-saline condition but not during cerebellar inactivation

- Task paradigm: If the monkey got a trial wrong, the same symbol was presented again in the next trial (repeat condition); else, one of the two symbols was presented with equal probability in the next trial.
- Beeswarm plot of the probability of repetition of the same symbol after a wrong trial, in the normal task (left) and the repeat task (right). $P = 1.9603 \times 10^{-8}$; ranksum test.
- Average learning curve (from both monkeys) for saline-control (S), saline-control with repeat (S(r)), Muscimol (M) and Muscimol with repeat (M(r)) conditions.
- Quantitation from **Fig S3c** in the trial window of 10-20 trials, for individual sessions, for two monkeys separately (circles are Monkey B; squares are Monkey S). Repetition of symbols during learning significantly improved the performance (rate of learning) during control-saline condition ($P = 0.0029$; paired t-test) but not during cerebellar inactivation ($P = 0.1117$; paired t-test).

Figure S4: Performance within a day for different learning sessions

- Behavioral performance for monkey B during saline-control condition across sessions for each day, for OT, beginning of learning (L_{beg}) and end of learning (L_{end}).
- Same as **Fig S4a**, but during cerebellar inactivation condition.
- Same as **Fig S4a**, but for monkey S.
- Same as **Fig S4b**, but for monkey S.

Figure S5: Hand positions for both monkeys for bar release condition

Horizontal (H) and vertical (V) hand positions triggered on the movement onset, for Monkey B (left) and Monkey S (right) during OT task for bar release movement.

Figure S6: Cerebellar inactivation did not impair the ability to make different hand movements

- Top left: Hand position in space and time (color indicates the time as in the color bar inset) for a representative trial during saline condition. Right: Snapshots from high-frame rate movies showing the monkey's hand movement trajectory at four time points (0, 200 and 450 and 700 ms from the start of movement) for bar release. Red circle highlights the fluorescent marker. Bottom: same as top, but for a representative trial during Muscimol condition.
- Same as a, but for dowel release condition.
- Left: Average horizontal (H) and vertical (V) hand trajectories for bar release, aligned on movement onset during saline (grey) and Muscimol (yellow) conditions. Right: Same as left but for dowels condition.
- Behavioral performance of monkeys (monkey B: circles, monkey S: squares) during overtrained condition for bars vs dowels under Muscimol (left; $P = 0.3744$; paired t-test); and between dowels for saline and Muscimol (right; $P = 0.7009$; paired t-test).

Figure S7: Cerebellar inactivation did not have differential effects on motor parameters for correct and wrong trials during learning

- Amplitude of hand position on correct trials, OT, Monkey B (left): $P = 0.9849$; ranksum test; monkey S (right), $P = 0.5362$; paired t-test
- Amplitude of hand position on correct trials, learning, Monkey B (left): $P = 0.4614$; paired t-test; Monkey S (right): $P = 0.5307$; ranksum test
- Probability of licking on correct trials, OT, Monkey B (left): $P = 0.9176$; paired t-test; Monkey S (right): $P = 0.6318$; paired t-test
- Probability of licking on correct trials, learning, Monkey B (left): $P = 0.6636$; t-test; Monkey S (right): $P = 0.7334$; paired t-test
- Visual exploration on correct trials, OT, Monkey B: $P = 0.9527$; ranksum test; Monkey S: $P = 0.9999$; ranksum test
- Visual exploration on correct trials, learning, Monkey B: $P = 0.9098$; ranksum test; Monkey S: $P = 0.4624$; paired t-test
- Hand position on wrong trials, OT, Monkey B (left): $P = 0.6509$; ranksum test; Monkey S (right): $P = 0.5833$; paired t-test
- Hand position on wrong trials, learning, Monkey B (left): $P = 0.1989$; ranksum test; Monkey S (right): $P = 0.5877$; ranksum test.
- Probability of licking on wrong trials, OT, Monkey B (left): $P = 0.6208$; ranksum test; Monkey S (right): $P = 0.9908$; t-test
- Probability of licking on wrong trials, learning, Monkey B (left): $P = 0.2711$; t-test; Monkey S (right): $P = 0.8054$; paired t-test.
- Visual exploration on wrong trials, OT, Monkey B: $P = 0.5932$; ranksum test; Monkey S: $P = 0.6048$; ranksum test
- Visual exploration on wrong trials, learning, Monkey B: $P = 0.2424$; ranksum test; Monkey S: $P = 0.1058$; paired t-test.

Figure S8: Cerebellar inactivation perturbed the win stay lose switch strategy

- Reinforcement-drift diffusion model. Action choices are modeled as accumulators with rates v_{a1} and v_{a2} , racing to threshold (bound). The winner takes all and consequences of the chosen action a_{ch} is evaluated by the activity of P-cells given by Δ . This is used to update the rates of the accumulator on a trial by trial basis. For saline-control condition, Δ is given by the difference in firing rate of correct and wrong outcome in the delta epoch of the P-cells (Δr). For muscimol condition, this is a random number drawn from a Gaussian distribution.
- Probabilities of win-stay and lose-switch during learning from both monkeys during saline-control condition
- Same as **Fig S8a** but during cerebellar inactivation condition.
- Reinforcement learning model prediction for saline-control condition with intact P-cell delta activity.
- Reinforcement learning model prediction for Cerebellar inactivation condition with absent P-cell delta activity.

Figure S9: Inactivation of anterior cerebellum did not affect monkey's performance

- T1 MRI showing injection regions in the mid-lateral cerebellum (yellow) and anterior cerebellum (green) for monkey B. Same format as **Fig S1**.
- Same as above but for monkey S.
- Behavioral performance of both nonkeys during OT and learning trials for control-saline (gray) and anterior cerebellar inactivation (green) conditions.
- Quantitation of mean squared error ($MSE = \text{bias}^2 + \text{variance}$) from **Fig S9c**;
- RT vs % error for saline injections (S; gray), muscimol injections in the mid-lateral cerebellum (M_{ml}; yellow) and anterior cerebellum (M_a; green). Each maker is an individual session (for one repetition of association learning). Triangles with error bars are means and s.e.m.
- The % error between S and M_{ml} was significant (1.1441×10^{-10} ; ranksum test) and so was between M_a and M_{ml} (2.0153×10^{-9} ; ranksum test) but not between S and M_a (0.1217 ; ranksum test)
- The RT between S and M_{ml} was significant (1.5543×10^{-4} ; ranksum test) and so was between M_a and M_{ml} (0.0159 ; ranksum test) but not between S and M_a (0.8732 ; ranksum test)

Methods:

Animal subjects and surgery

We have described all procedures in detail elsewhere⁴. We describe them here briefly. We used two male adult rhesus monkeys (*Macaca mulatta*), B and S, for all the experiments. All experimental protocols were approved by the Animal Care and Use Committees at Columbia University and the New York State Psychiatric Institute, and complied with the guidelines established by the Public Health Service Guide for the Care and Use of Laboratory Animals. We located the cerebellum in each monkey using a Simons 3T magnet. We localized the mid-lateral cerebellum using Brainsight, Rouge research Inc. and accessed them through two recording cylinders, on the left hemisphere of monkey B and right hemisphere of monkey S.

Tasks:

We used the NIH REX system for behavioral control²⁸. The monkey sat inside a dimly lit recording booth, with its head firmly fixed, in a Crist primate chair 57 mm in front of a LCD monitor upon which visual images were presented which were controlled by a PC running the NIH VEX graphic system.

Two-alternative forced-choice discrimination task: The task began with the monkeys grasping two bars, one with each hand, after which a white 1° x 1° square appeared as a trial cue for 800 ms. Then one of a pair of symbols appeared, briefly for 100 ms in some sessions or until the monkey initiated a hand response in other sessions, at the center of gaze. One symbol signaled the monkey to release the left bar and the other to release the right bar. We rewarded the monkeys with a drop of juice for releasing the hand associated with that symbol. The liquid reward and a beep paired with opening of the solenoid were delivered immediately (with a delay of 1 ms) after the initiation of the correct hand movement. We trained the monkeys to associate a specific pair of symbols (green square and pink square) with specific actions (left and right-hand release, respectively) for about 4-6 months until their performance was above 95% correct; we refer to this as the overtrained condition.

In the visuomotor associative learning version of the task (**Fig 1a**), we began every recording session by presenting the monkeys with the same overtrained symbol pair (overtrained condition), and after a number of trials, switched the symbol pair to two fractal symbols, which the monkey had never seen before (novel condition). Both the symbols had equal probabilities of occurring in every trial. The manipulanda remained the same throughout the task. We used two versions of the visuomotor associative learning task shown in **Fig S3**. In the more difficult version, the symbols appeared with equal probability regardless of the monkey's prior decision. In the easier version, we used an error-correction strategy, showing the same symbol until the monkey made a correct decision. Monkeys learned the symbol-hand association faster during this paradigm, under control conditions. The manipulanda remained the same throughout the task.

In the manipulanda change task (**Fig S6**), we began every recording session by presenting the monkeys with the same over-trained symbol pair and bar manipulanda, and after a number of trials, switched the bar manipulanda to dowel manipulanda. The visuomotor association remained the same

Data collection and analysis

Infusion protocol: At the beginning of each recording session, we lowered a cannula filled with 10 μ L of either saline or 10 μ g/ μ L muscimol (diluted in 1X phosphate buffered saline solution), in a track selected on the basis of previous recordings near crus I and II of the mid-lateral cerebellum. We slowly infused the solution using a syringe pump (Harvard Apparatus) through a direct line to the cannula at a constant rate (0.2-0.5 μ L/min for 20-30 min). Therefore, we delivered a total mass of 4-10 μ g of muscimol in each session. Previous studies have reported that muscimol diffuses and hence functionally inactivates the neurons in an estimated span of 2-3 mm of spherical radius^{29,30}. Infusions were typically made at different depths within a single track (total span of ~ 5 mm) to increase the diffusion radius of the chemical. After infusion, the cannula was left for at least 20-30 min *in situ*. Occasionally, we lowered an electrode in the same track at the same depth after the chemical infusion to confirm silencing of neurons.

Single unit recording: We introduced glass-coated tungsten electrodes with an impedance of 0.8-1.2 MOhms (FHC) into the left mid-lateral cerebellum of monkeys every day that we recorded

using a Hitachi microdrive. We passed the raw electrode signal through a FHC Neurocraft head stage, and amplifier, and filtered through a Krohn-Hite filter (bandpass: lowpass 300 Hz to highpass 10 kHz Butterworth), then through a Micro 1401 system, CED electronics. We used the NEI REX-VEX system coupled with Spike2 (CED electronics) for event and neural data acquisition. We verified all recordings off-line to ensure that we had isolated Purkinje cells⁴.

Analysis of the learning behavior: We constructed the learning curve for every session by calculating the percent correct trials in a sliding window of 10 trials as the bin width moved by 5 trials. Then, we averaged the learning curves across repeated sessions for each symbol pair for each monkey (16 for monkey B and 9 for monkey S) to obtain a total of 25 learning curves. We estimated when the performance first increased above 75% and referred to it as the acquisition-difficulty level for each session. We repeated this for both saline-control condition and the muscimol-inactivation condition.

Hand tracking: We painted a spot on the monkeys' right hand with a UV-blacklight reactive paint (Neon Glow Blacklight Body Paint) prior to every session. We used a 5W DC converted UV black light illuminator to shine light on the spot. Then we used a high speed (250 fps) camera (Edmund Optics), mechanically fixed to the primate chair, to capture a video sequence of the hand movement while the monkeys performed the tasks. We used the track mate feature^{31,32} and custom written software in MATLAB to semi-manually track the fluorescent paint spot painted on the monkey's hand⁴.

Licking: We recorded licking at a sampling rate of 1000 Hz using a capacitive touch sensor coupled to the metal water spout which delivered liquid water reward near the monkey's mouth. Raw binary lick traces were used to generate instantaneous lick rate by trial averaging and smoothing it with a Gaussian kernel of sigma = 20 ms.

Eye movements: We tracked the monkey's left eye positions at 240 Hz sampling rate with an infrared pupil tracker (ISCAN, Woburn, MA USA) interfaced with Spike 2 (CED electronics) where it was upsampled to 1000 Hz and synced with the event markers from NEI REX-VEX

system. To calculate the radius of visual exploration (r_{exp}), we first create a heatmap with overlaid eye movements (**Fig 3e-f**). We then fit a 2D gaussian distribution to this data to see the extent of spread from the center and calculated the full-width-at-half-maximum.

Reinforcement-drift diffusion model: Details of the model are described elsewhere⁴. We briefly describe it here. Consider learning one symbol-action-outcome association for one symbol with two alternative action choices, a_L and a_R corresponding to left and right hand bar-release hand movements respectively. We model the action selection through a race to threshold model where there is a race in the evidence accumulation between the action values a_L and a_R modeled by Wiener first-passage time (WFPT) distribution. This calculates the likelihood of the reaction time of choice i given by:

$$V_i(x) \sim WFPT [v_i(t)]$$

where $v(t)$ denotes the rate of accumulation process for the trial t . The rate of accumulation in the overtrained (OT) condition, v_i^{OT} is assumed to be a constant. Then, we model the evolution of $v_i(t)$, with learning as:

$$v_i(t+1) = v_i(t) + \left(\frac{m}{r}\right) \Delta_i(t) I_c(t) I_i(t)$$

where

$$\Delta_i(t) = r [v_i^{OT} - v_i(t)] I_c(t) I_i(t) + (I_w(t) + I_c(t) I_{i \neq j}(t)) \Delta_i(t-1)$$

here, m captures the rate of learning, r is the scale factor, $I_i(t)$ is the indicator function describing the chosen action (a_{ch}) on trial t defined as follows:

$$I_i(t) = \begin{cases} 1 & \text{if } a_{ch}(t) = a_i \\ 0 & \text{if } a_{ch}(t) \neq a_i \end{cases}$$

We propose that the rate of change of the accumulation rate $v_i(t)$, given by $\Delta_i(t)$ is represented in the delta epoch of the P-cell population that fired at the time of interest. We estimated the accumulator's initial and final rates by fitting a drift-diffusion model to the observed RT values. Then, we minimized the error between the generated and the empirical RT distributions for the OT and initial learning conditions to get estimates of the parameters. We then simulated the learning process with these values and calculated the probability of win-stay and lose-switch

as a function of learning (**Fig S8**). For cerebellar inactivation condition, we modeled the firing rate of P-cells in the delta epoch, $\Delta_i(t)$ to be a random Gaussian noise and repeated all the above steps otherwise identically (**Fig S8**).

Statistics: To check if two independent distributions were significantly different from each other, we first performed a two-sided goodness of fit Lilliefors test, to test for normality, then used an appropriate t-test; or else a non-parametric Wilcoxon ranksum test. All error bars and shading in this study, unless stated otherwise, are mean $\pm$ s.e.m.