



No Evidence for Semantic Prediction Deficits in Individuals with Cerebellar Degeneration

Maedbh King^{1}, Sienna Bruinsma^{1*}, Richard B. Ivry^{1,2}*

1. Department of Psychology, University of California, Berkeley, CA, USA

2. Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

* These authors contributed equally.

+ To whom correspondence should be addressed: Maedbh King, Department of Psychology, University of California, Berkeley, 2121 Berkeley Way West, Berkeley, CA, USA. Email:

maedbhking@berkeley.edu

Contributions: M.K. and R.B.I. originally conceived of the project. M.K., S.B., and R.B.I. designed the experiment. S.B. collected the data. M.K. and S.B. performed the analyses. M.K., S.B., and R.B.I. wrote the manuscript. All authors discussed the results and contributed to the final manuscript.

Keywords: cerebellum; semantic prediction; cerebellar degeneration; internal model; prediction

Acknowledgements: This work was supported by the National Institutes of Health (NS092079, NS105839 to R.B.I). Thanks to Jonathan Tsay and Amanda LeBel for providing helpful feedback on the manuscript.

Competing Interests: R.I. is a co-founder with equity in Magnetic Tides, Inc. The other authors declare no competing interests.

Abstract

Cerebellar involvement in language processing has received considerable attention in the neuroimaging and neuropsychology literatures. Building off the motor control literature, one account of this involvement centers on the idea of internal models. In the context of language, this hypothesis suggests that the cerebellum is essential for building semantic models that, in concert with the cerebral cortex, help anticipate or predict linguistic input. To date, supportive evidence has primarily come from neuroimaging studies showing that cerebellar activation increases in contexts in which semantic predictions are generated and violated. Taking a neuropsychological approach, we put the internal model hypothesis to the test, asking if individuals with cerebellar degeneration ($n=14$) show reduced sensitivity to semantic prediction. Using a sentence verification task, we compare reaction time to sentences that vary in terms of cloze probability. We also evaluated a more constrained variant of the prediction hypothesis, asking if the cerebellum facilitates the generation of semantic predictions when the content of a sentence refers to a dynamic rather than static mental transformation. The results failed to support either hypothesis: Compared to matched control participants ($n=17$), individuals with cerebellar degeneration showed a similar reduction in reaction time (RT) for sentences with high cloze probability and no selective impairment in predictions involving dynamic transformations. These results challenge current theorizing about the role of the cerebellum in language processing, pointing to a misalignment between neuroimaging and neuropsychology research on this topic.

Introduction

Anatomical, neuropsychological, and neuroimaging work over the last 35 years has implicated the cerebellum in functions extending well beyond the motor domain. Throwing down the gauntlet for a “cognitive cerebellar revolution”, Leiner, Leiner, & Dow (1986) highlighted the expansion of the cerebellum across vertebrate evolution and in particular, a parallel expansion of the neocerebellum and prefrontal cortex. They hypothesized that this pattern likely reflected extensive communication between the cerebellum and cortical association regions, connections that could support mental coordination in a manner analogous to how cerebello-cortical connectivity supported motor coordination. Over the following decades, the neuropsychological, neuroimaging, and brain stimulation literatures have provided ample evidence in line with this general perspective. Behavioral impairments on a variety of cognitive and affective tasks have been described in individuals with cerebellar dysfunction (Schmahmann & Sherman, 1998; Ivry & Fiez 2000; Sokolov et al. 2017; Kansal et al. 2017) and activation is consistently observed in the cerebellum during a broad range of tasks that is not readily related to motor preparation or production (Stoodley et al. 2009; King et al. 2019).

One prominent domain of interest is language. In their seminal PET study, Petersen et al. (1989) used a subtractive logic to identify brain regions involved with different linguistic processes. Their results revealed prominent activation of the right cerebellum in a contrast designed to show regions involved in semantic retrieval while controlling for articulatory preparation. This observation has been confirmed in many subsequent neuroimaging studies investigating semantic retrieval and prediction, with the activation centered in right medial-to-lateral Crus I/II (Fiez et al., 1996; Moberget et al., 2016; Lesage et al., 2017). The neuropsychology literature presents a more problematic picture. While speech dysarthria is a prominent feature in some forms of cerebellar ataxia (Ackerman & Hertrich, 1994; Richter et al., 2007), the evidence is mixed in terms of whether cerebellar dysfunction, at least when emerging in adulthood, disrupts semantic processing. In a widely cited case study by Fiez et al. (1992), a patient with a right inferior cerebellar lesion had marked difficulty on a semantic generation task, either coming up with an inappropriate semantic associate to a target noun or failing to follow instructions and simply repeating the target word. However, subsequent studies failed to find group-level differences on similar tasks (Helmuth et al. 1997; Gasparini et al., 1999; Riva, 1998; Richter et al., 2004). While it is possible that these null results reflect the heterogeneity in the patient samples, it may be that semantic generation

tasks are not sensitive to cerebellar-related language impairments or optimal for assessing computations performed by the cerebellum in semantic processing.

Theoretically, researchers have turned to the sensorimotor control literature in considering how the cerebellum may contribute to semantic processing. One key idea has centered on the notion of prediction and, specifically, how the cerebellum may generate internal models to anticipate future input based on the current context (Wolpert et al. 1998; Ito, 2008). In the motor context, an internal model may represent the dynamic properties of an object (e.g., body part, tool), allowing the agent to anticipate the sensory consequences of an action within a particular environment (Wolpert et al., 1995). Extending this idea to language, an internal model would operate on abstract semantic knowledge and input from the current discourse to generate semantic expectancies; for example, to anticipate the speaker's next utterance or anticipate when the speaker would pause to facilitate fluid turn taking (Moberget et al., 2014; Lesage et al., 2012, 2017).

Semantic verification tasks have provided one testbed of the internal model hypothesis. In an fMRI study, Moberget et al. (2014) compared the hemodynamic response to three types of sentences: 1) Congruent in which the last word was highly predictable given the preceding, base phrase; 2) Incongruent in which the last word violated a prediction established by the base phrase; 3) Scrambled in which the words in the base phrase were randomly shuffled, thus precluding the generation of a prediction. An analysis time-locked to the appearance of the last word revealed two key cerebellar results. First, violations of predictions produced bilateral activation encompassing posterolateral Crus I/II compared to either congruent or scrambled sentences. Second, compared to scrambled sentences, congruent sentences produced activation in a more circumscribed portion of posterolateral cerebellum, now restricted to the right cerebellar hemisphere only. Thus, these data are in accord with the hypothesis that the cerebellum not only processes violations of predictions (which might suggest a role in error-detection/error-correction), but also is engaged in the generation of predictions. Lesage et al. (2017) reported similar effects in an event-related fMRI design that manipulated sentence predictability. Not only did they find that activity in the right posterolateral cerebellum correlated with the predictability of the target word, but they also demonstrated that this same region was engaged in phonological, but not semantic or orthographic processing.

Behavioral assays of cerebellar involvement in language have come from two sources, non-invasive brain stimulation targeted at the cerebellum in healthy individuals (Argyropoulos 2016; Lesage et al. 2012; D’Mello et al. 2017; Gatti et al. 2021) or experiments involving individuals with cerebellar pathology (Stoodley & Schmahmann, 2009; Alexander et al. 2012; Kansal et al. 2017). Lesage et al. (2012) instructed participants to shift their gaze to a peripheral object that was predicted by the semantic content of a spoken sentence. After rTMS was applied targeting right posterolateral cerebellum, participants took longer to fixate on the predicted object. In contrast, rTMS had no effect when the semantic content did not predict the target object and the selective effect on predictive looking was not observed when the rTMS was directed at the vertex, a control stimulation location.

Along the same lines, D’Mello et al. (2017) showed that, after tDCS over the right lateral cerebellum, the BOLD response in the cerebellum was selectively elevated in response to predictive sentences. Interestingly, the change in activation was not accompanied by a change in behavior. Participants were faster to verify predictable sentences compared to non-predictive, and this advantage was similar in the tDCS and control conditions. Similarly, Moberget et al. (2016) failed to observe any difference in behavior between individuals with cerebellar pathology and matched control participants. Thus, there is a mismatch between the imaging/neurostimulation and neuropsychology results. Consistent with the internal model hypothesis, increased activation is observed in the right posterolateral cerebellum when the semantic content of a sentence supports the formation of a prediction. However, the evidence is mixed in terms of whether lesions encompassing this region disrupts the ability to make semantic-based predictions. More generally, the neuropsychological literature fails to provide strong evidence of a role of the cerebellum in semantic processing. While individuals with cerebellar degeneration consistently show impairment in tests of fluency or word completion (Stoodley & Schmahmann, 2009; Alexander et al. 2012; Kansal et al. 2017), these tasks heavily tax processes associated with cognitive control.

In the present study, we set out to conduct a new neuropsychological test of the internal model hypothesis, comparing the performance of individuals with cerebellar degeneration (CD) and matched controls on a sentence verification task. Similar to our earlier studies (Moberget et al., 2014; 2016), participants made a two-alternative forced choice response on a semantic prediction task, indicating if each sentence was meaningful or meaningless. We emphasized reaction time in the current study, under the assumption that this would provide a more sensitive measure of performance than accuracy. We

manipulated the strength of the prediction of the final word in the sentence given its base phrase. To this end, we manipulated cloze probability, a common linguistic metric that quantifies the context-dependent predictability of a word. Our main analysis focused on the meaningful sentences. We expected that control participants would have faster reaction times in response to target words with high cloze probability compared to target words with low cloze probability (Staub et al. 2015), reflecting the benefits of predictive processing when the base phrase allows the participant to anticipate the target word. If the integrity of the cerebellum is important for generating semantic predictions, we expect that the CD group would show an attenuated effect of cloze probability: Specifically, the CD group should show a reduced benefit to target words with high cloze probability.

As a secondary test of the prediction hypothesis, we analyzed reaction times for the meaningless sentences to examine how the two groups responded to violations of semantic predictions. This analysis is problematic for two reasons. First, these sentences entail two types of prediction violations. There is the general semantic violation given that the final word results in a nonsensical sentence. And, at least for the high cloze sentences, there is a prediction violation in that the final word does not match the expected word. Second, we did not have a strong *a priori* hypothesis here, even for the controls. Reaction times may be faster when the base phrase suggests a strong prediction since the violation would be easier to detect. Alternatively, reaction times might be slower when the base phrase suggests a strong prediction since anticipation of the expected final word might cause interference in processing the target word. Nonetheless, the meaningless sentences provide a test bed to see if the CD and control groups showed different response time patterns to semantic prediction violations.

We also designed the experiment to test a second hypothesis concerning predictive functions of the cerebellum. Even in the motor domain, prediction is not the exclusive domain of the cerebellum; indeed, one could argue that most neural activity is fundamentally about prediction (Friston, 2009). This leads to the question of what might be constraints on cerebellar-dependent predictions? Various candidates have been proposed over the years, some that specify the task (e.g., sensory consequences of movements generated by the skeletal musculature Day et al. 1998) or computational (e.g., timing, error-based learning, Keele & Ivry 1990; Wolpert et al. 1998) domain. We have recently proposed that the cerebellum is essential for predictions that require a continuous representational transformation (CoRT, McDougle et al. 2022). The key idea is that the cerebellum facilitates mental processing when an

internal representation undergoes a continuous, dynamic transformation, but is not essential when the mental operations entail more discrete or transformations or maintenance of static representations. In our initial work on this problem, the CoRT constraint provided a parsimonious account of the pattern of impairments (and spared performance) in the domains of visual cognition and arithmetic. For example, individuals with CD were impaired in the rate at which they mentally manipulated a visual image (Shepard & Metzler, 1971), whereas they showed a normal processing rate in the iterative evaluation of a series of discrete spatial representations.

In the context of language, we hypothesize that a CoRT-like operation would be relevant when the semantic content of a sentence conveys a dynamic representational transformation. For example, the sentence “The man cut the rope with a pair of scissors” evokes a strong, dynamic mental simulation, with a conceptualization of the context (an agent, desiring to cut a rope), helping anticipate the final word. In contrast, the sentence “Spring was her favorite season of the year” is relatively static. One could say it requires an internal model, one that captures the idea that a person has their idiosyncratic preferences for a season. But this model does not evoke a dynamic simulation. Based on this reasoning, we also designed the experiment such that we could compare sentences that described dynamic transformations (CoRT-dependent) with those that described static situations (non-CoRT). Although we do not have an *a priori* prediction for the control participants on this dimension, we predicted that the CD group would be relatively impaired on the sentence verification task for the dynamic sentences.

Methods

We tested individuals with cerebellar degeneration (CD) and matched controls (CO) on a semantic verification task. Participants were asked to determine whether a sentence was meaningful or meaningless based on the final word of the sentence. We manipulated two variables, prediction strength and dynamic context in a 2 x 2 design (Fig 1). All testing, including the neurological evaluation of the CD participants was conducted on-line.

We first describe the procedures involved in creating the stimulus materials for the study and then turn to the procedure for the main experiment. The study was approved by the institutional review board at the University of California, Berkeley.

Construction and Evaluation of Stimuli

Sentences were drawn from two sentence databases, both of which included cloze probability ratings (Block & Baldwin, 2010; Peelle et al. 2020). Together, there are 3,583 sentences in the two databases. The 498 sentences in the Block and Baldwin database have a minimum cloze probability of 0.3. A cloze probability is calculated as the proportion of participants who give the same response to an incomplete sentence or passage (Kutas & Hillyard, 1984). We applied this same minimum criterion to the Peelle database which excluded 923 of the 3085 sentences in that database, yielding a total of 2660 sentences in the initial inclusion stimulus set.

We enlisted raters to evaluate the degree to which these 2660 sentences conveyed a dynamic event (“CoRT-ness rating”). We initially recruited undergraduate participants (n=8, 4 female/4 male) through a research participant pool at the University of California, Berkeley. These individuals completed the rating procedure in the lab. However, with the onset of institutional restrictions related to the COVID-19 pandemic, we transitioned to online methods. For these ratings, we recruited participants (n=61) through Prolific (<https://www.prolific.co/>), an online participant recruitment platform. We verified via self-report that the first language of all participants was English. Participants received monetary compensation for their time.

Participants were asked to rate each sentence in terms of how strongly the content evoked a dynamic event, using a 5-point Likert scale where 1 indicated “strongly disagree” (i.e., strongly non-CoRT) and 5 indicated “strongly agree” (i.e., strongly CoRT). Participants were provided examples of sentences conveying a strong dynamic event (e.g., “The man loosened the tie around his neck”) and a weak dynamic event (e.g., “Her job was easy most of the time.”) As a primer, participants were given a list of five sentences with the correct ratings to ensure that they understood the goal of the task.

Each trial began with the presentation of a fixation cross. After 500 ms, a target sentence was displayed at the center of the screen, along with the instruction prompt at the bottom of the screen, “Does this sentence evoke strong, changing imagery?” Participants had 10 seconds to indicate their response using the number keys “1” to “5”. The fixation cross appeared as soon as the rating was registered. The in-person participants rated all 500 of the sentences in the Block & Baldwin dataset. The stimulus order was randomized across individuals and breaks were provided every 180 trials. For the online

participants, we used the Gorilla platform (<https://gorilla.sc/>) to run the experiment. The 2160 sentences retained from the Peelle database were randomly divided into 12 sets of 180 sentences each. Each participant rated one set, with a break provided after 90 trials. The stimuli within a set were randomized across individuals.

For the final stimulus set, we used the cloze probability information and CoRT ratings to select 125 sentences for each of the four cells created by the factorial crossing of prediction and dynamics (Fig 1 A). On the prediction dimension, sentences had to have a cloze probability between 0.3 and 0.5 to be selected as weak exemplars, or between 0.8 and 1.0 to be selected as strong exemplars. In this way, we created a similar cloze range for each level. For the dynamics dimension, sentences had to have a mean rating of less than 2 (static) or mean rating greater than 4 (dynamic). For high cloze sentences, dynamic conditions had an average CoRT rating of 4.214 (std=.24), and static conditions had a rating of 1.616 (std=.306). The average values of high cloze probabilities were roughly equivalent across dynamic (mean=.885 std=.053) and static (mean=.878, std=.049) ratings. For low cloze sentences, dynamic conditions had an average CoRT rating of 4.357 (std=.298), and static conditions had an average rating of 1.655 (std=.305). The average values of low cloze probabilities were roughly equivalent across dynamic (mean=.407, std=.058) and static (mean=.398, std=.057) ratings.

As a final step, we conducted a pilot of the semantic verification task, with participants making speeded responses to indicate if the sentence was meaningful or non-meaningful (detailed below). Participants (n=77) were recruited via Prolific and tested across 6 blocks with the full set of 500 sentences, with a break provided approximately every 80 trials. Based on the results of the pilot study, we designed the main experiment to include 5 blocks each comprising 64 trials. We opted to reduce the number of blocks from 6 (in the pilot) to 5 (main experiment) to reduce the experiment time for our patient group. We divided the 500 pilot sentences into four conditions: dynamic, high cloze; dynamic, low cloze; static, high cloze; static, low cloze. Within each of those groups, we chose 80 sentences that had the lowest standard deviation across the pilot participants, resulting in a total of 320 sentences for the main experiment.

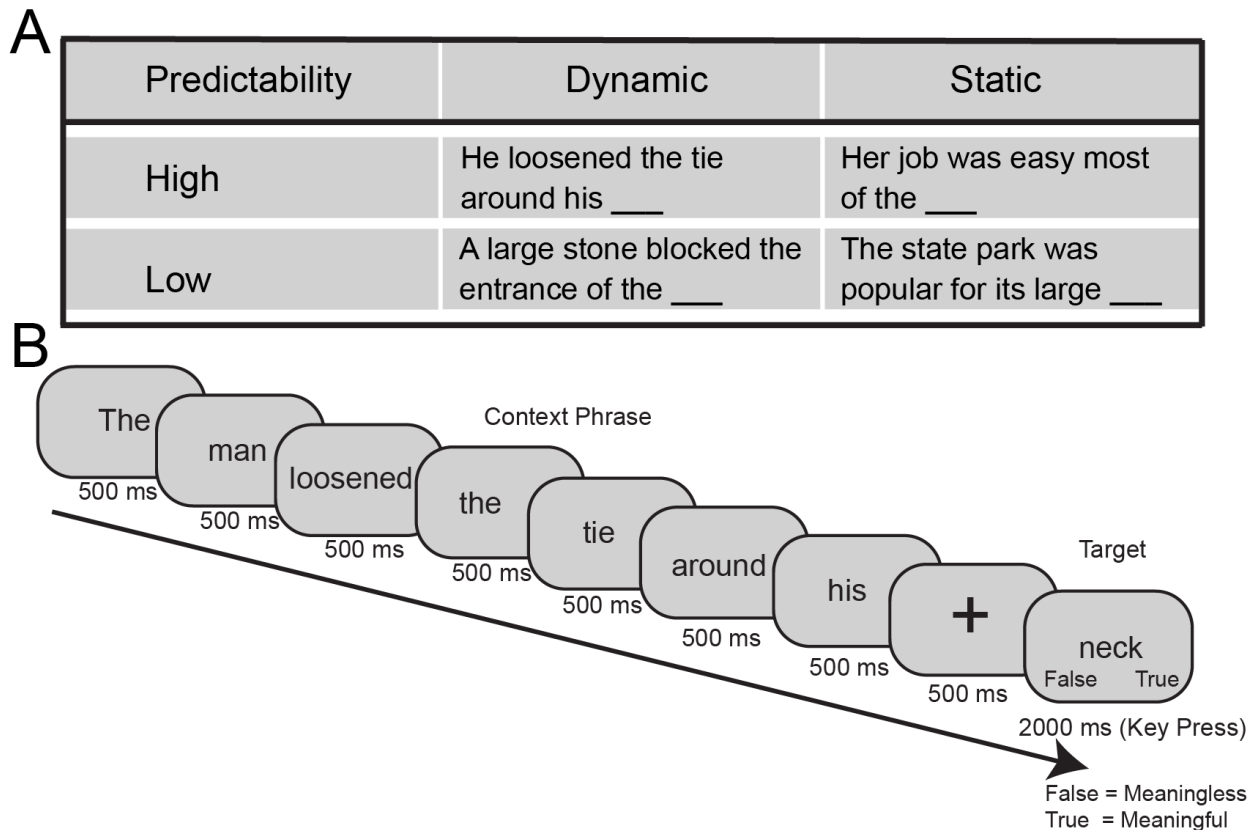


Fig 1: Task Design. A) Representative sentences for evaluating the internal and CoRT models. Sentence predictability is manipulated via high and low cloze probabilities, and representational transformations are manipulated via dynamic and static simulations. B) Example trial. Each word of the stem (“the man loosened the tie around his”) is presented for 500 ms, followed by a fixation cross (500 ms). Participants then have 2000 ms to make a key press indicating “true” or “false” on the target word (“neck”). True indicates a meaningful sentence, while false indicates a meaningless sentence.

Main Experiment

Participants

17 individuals with cerebellar degeneration (CD) and 19 age- and education-matched control participants (CO) were recruited, drawing on a patient database maintained by our lab. The database includes individuals recruited from ataxia support groups around the country and through advertisements posted on the National Ataxia Foundation webpage. Researchers from our group send recruitment letters to support group leaders and these leaders share our letters with their members. Interested individuals contacted the lab and after completing a series of screening procedures, are entered into the CD database

if they meet inclusion criteria (Saban & Ivry, 2021). We then invited these individuals to participate in our experiment. Control participants were recruited via advertisements circulated on internet forums (e.g., craigslist). Although we have limited demographic data on the CO group, we assume they are selected from a similar geographic distribution as the CD group given that the only restriction was that they reside in the United States. Participants were compensated with either \$20 or a \$20 Amazon gift card for their participation.

Table 1 provides demographic, diagnostic, neurological, and neuropsychological information for each individual in the CD group, along with summary information for the CO group. For the purposes of this study, we included individuals who had a diagnosis of cerebellar ataxia based on clinical exam (and MRI evidence as indicated by self-report) or diagnosis of spinocerebellar ataxia (SCA) based on genetic testing and/or family history. We excluded individuals with Friedrich's ataxia. For the neurological exam of ataxia, we used the Scale for the Assessment and Rating of Ataxia (SARA) which has a total possible score of 40 (maximum impairment). The test was modified for on-line administration. Specifically, items assessing gait, stance, and sitting were scored using a flow chart of questions rather than through direct observation given concerns about remote testing. To establish general neuropsychological status, we used the Montreal Cognitive Assessment (MoCA). For on-line testing, we eliminated the trail making item. Although the maximum score on the modified MoCA is 29, the scores reported below were adjusted so that they fall on a 30-point scale. Participants in the CO group were also administered the MoCA.

Three individuals in the CD group (p06, p08, p11) and one control (c05) were not included in the final analysis because of poor performance on the sentence verification task (overall accuracy below 70% on meaningful and meaningless sentences). Although this criterion may eliminate individuals with the most severe language problems, we were concerned it was more reflective of disengagement with the task or failure to understand the task instructions. Indeed, the mean accuracy score for these three was well below that of any of the other participants, with two performing poorly on the meaningful sentences and one responding "meaningful" to all of the meaningless sentences. For the remaining 14 individuals in the CD group, the mean score on the SARA test was 10.64 (SD = 5.36), a value that corresponds to mild-to-moderate impairment. At the time of testing, an average of 6.8 years (std=7.6) had elapsed since

diagnosis of cerebellar ataxia, and the participants reported symptom onset preceded diagnosis by about 4 years.

We selected individuals for the CO group by referencing our database to produce matches in terms of age and years of education (yoe). As can be seen in Table 1, the group means were comparable for the CD (mean age = 60.5; mean yoe = 16.63) and CO (mean age = 52.44; mean yoe = 16.25) groups, and there was no statistically significant difference between the groups for either years of education ($t_{30} = .34$, $p = .74$) or age ($t_{30} = 1.733$, $p = 0.09$).

Although not part of our recruitment criteria, there was no significant difference between the CD (mean = 25.93, std = 2.83) and CO (mean = 26.39, std = 2.72) groups on the MoCA test ($t_{30} = 0.49$, $p = 0.63$). Note that a MoCA score below 26 is indicative of possible cognitive impairment. Three of the CD participants in the final sample scored below this criterion as did 6 of the CO participants. Given our focus on one important aspect of cognition, language, we did not exclude participants based on their MoCA score. Moreover, there was no relationship between the overall MoCA score and two performance metrics: task accuracy ($R = 0.28$, $p = 0.12$) and reaction time ($R = 0.01$, $p = 0.59$).

ID	Hand	Age	Sex	MoCA	SARA	Years of Education	Etiology	Included
p01	R	48	F	29	4	25	SCA-1	Yes
p02	L	55	F	27	10	18	SAOA	Yes
p03	N/A	48	F	N/A	N/A	22	SCA-3	Yes
p04	R	51	F	26	2	12	SCA-6	Yes
p05	R	39	F	27	13	16	SAOA	Yes
p06	R	89	F	19	15	12	SCA-3	No
p07	R	80	F	26	18.5	12	SAOA	Yes
p08	R	79	M	22	8	16	SCA-6	No
p09	R	57	M	29	6	20	SCA-3	Yes
p10	R	42	F	29	10	14	SAOA	Yes
p11	R	67	M	26	15	20	SCA-5	No
p12	R	42	F	28	20	19	SAOA	Yes
p13	R	48	F	27	10	20	SAOA	Yes
p14	R	81	M	24	10	16	SAOA	Yes
p15	R	61	F	22	7	12	SCA-6	Yes
p16	L	53	F	27	17	18	SAOA	Yes
p17	R	59	F	27	9	12	SCA-6	Yes
control (mean)	—	52.44	—	26.39	—	16.25	—	—
control (std)	—	11.86	—	2.72	—	2.54	—	—
patient (mean)	—	60.5	—	25.93	10.64	16.63	—	—
patient (std)	—	15.31	—	2.83	5.36	3.86	—	—

Table 1: Demographic and clinical assessment for CD patients (p01-p17) and healthy controls (CO). MoCA scores are out of a total of 29 points (perfect score). SARA scores are out of a total of 40 points (most severe ataxia). CD=cerebellar degeneration; SAOA=spontaneous adult-onset ataxia. Note that participants p06, p08 and p11 were excluded from the final sample due to their poor performance on the semantic evaluation task.

Procedure

Participants were recruited via an email sent to individuals in our participant database. The email provided a brief overview of the experimental goal and requirements (e.g., English as first language, need to complete the 2-minute (62 trial) experiment in a single session). A link was provided to the experiment and participants were informed that they could initiate testing whenever they were ready since the testing was done in an automated manner. Each participant was provided with a unique link, the code that allowed us to match the participant with their experimental data.

The link took the participant to the Gorilla platform where the participant worked through a series of introductory screens that culminated in a page where they could provide electronic informed consent. Once this was obtained, the CD and CO participants completed a 2-minute task designed to provide a brief assessment of American English fluency. We modified the English LexTALE task (Lemhöfer & Broersma, 2012) for our online platform. On each trial, a fixation screen was presented for 250 ms, followed by a letter string. The participant indicated if the letter string formed an English word (press “J” on keyboard) or did not form an English word (press “F”). If a response was not entered within 2000 ms, the trial was recorded as “incorrect” and the program moved to the next trial. There were 62 trials, each with a unique target word (32 words, 32 non-words). All participants performed at a high level (overall mean = 96%, range = 70% - 100%).

Following the fluency assessment, the program moved on to the semantic verification task. The general procedure was similar to that employed in Moberget et al. (2014). The start of each trial was marked by the presentation of a fixation screen for 500 ms (Fig 1B). This was followed by the serial presentation of the sentence stem (e.g., “The man loosened the tie around his”), with each word displayed on the screen for 500 ms. We opted to use a serial presentation mode to minimize demands on eye movements.

Following the presentation of the stem, the fixation screen re-appeared for 500 ms, cueing the participant that the next screen would display the target word. The target word was then presented and the participant made a speeded response, indicating if the target word resulted in a meaningful sentence (e.g., “neck”) or non-meaningful sentence (e.g., “banana”). Responses were made by pressing the “J” key with the right index finger or “F” key with the left index finger for meaningful and non-meaningful sentences, respectively. The target word remained visible until a response was detected or 2 s elapsed. The screen then went blank for 500 ms before the onset of the next trial. Participants were instructed that

the primary dependent variable was response time; as such, while they should seek to achieve a high level of accuracy, they should respond as quickly as possible.

Participants first completed a practice block of 24 trials in which 16 of the sentences were meaningful and 8 were non-meaningful. This was followed by five test blocks of 64 trials each, or a total of 320 test trials. Across the 320 trials, there were 80 sentences for each of the four conditions (Table 1a), providing our assessment of prediction (high vs low cloze probability) and CoRT (dynamic vs static inference). Subject to this constraint, the stimulus set for each participant was based on random selection (without replacement) from the full set of 344 sentences, with the unselected items used in the practice trials. Within each of the four conditions, the target word was the final word for that sentence in the stimulus set on 70% of the trials (and, thus, formed a meaningful sentence). For the other 30% of the trials, the target word was determined by randomly generating a target word from all other target words. The experimenter then manually checked that the target word that was randomly chosen was subject to the constraint that it formed a non-meaningful sentence.

Each block took approximately 8 minutes and summary feedback in the form of percent correct for the previous block was provided at the end of each block, along with a reminder of the keyboard mappings. The participant moved on to the next block by pressing the spacebar. Upon completing the sentence verification task, the participant was asked to fill out a brief feedback form concerning their experience with the online testing platform. The entire experiment took approximately 50 minutes (longer if participants took longer breaks).

We did not include reaction times for the non-meaningful trials or for trials in which the response was incorrect. The mean values were then entered in a 2 (Group) x 2 (Prediction) x 2 (CoRT) within-subject ANOVA.

Results

As expected, mean accuracy was very high for both groups (CD (n=14): $M = .97$; $sd = 0.05$; CO (n=18): $M = .96$; $sd = 0.07$) across all four conditions (Fig 1A). While recognizing the limited value of an error analysis given that performance was near ceiling, we employed a 3-way within-subject ANOVA with the factors (Group) x 2 (Prediction) x 2 (CoRT). None of the main effects were significant (Group: $F_{1,120} =$

.04, $p = .85$, $\eta^2 < .001$; Prediction: $F_{1,120} = .46$, $p = .51$, $\eta^2 < .01$; CoRT: $F_{1,120} = .48$, $p = .49$, $\eta^2 < .01$). The Group factor did not interact with either Prediction ($F_{1,120} = 0.29$, $p = 0.59$, $\eta^2 < .001$) or CoRT ($F_{1,120} = 0.19$, $p = .66$, $\eta^2 < .001$), and the 3-way interaction was not significant ($F_{1,120} = .0001$, $p = 0.99$, $\eta^2 < .001$).

We analyzed the meaningful and meaningless trials separately in the analyses of the reaction time data, restricting these analyses to the trials in which the participants were accurate. For the meaningful sentences, the CD group was slower than the CO group by 206 ms (main effect of group: $F_{1,30} = 23.34$, $p < .001$, $\eta^2 = 0.44$). As can be seen in Figure 1B, participants became slightly faster over the course of the five test blocks (Fig. 1B; $F_{4,120} = 5.54$, $p < .001$, $\eta^2 = .16$), but the difference between the two groups was maintained. Given the absence of a Group x Block interaction ($F_{4,120} = 0.19$, $p = 0.94$, $\eta^2 < .001$), we collapsed the data across the five blocks in the subsequent analyses.

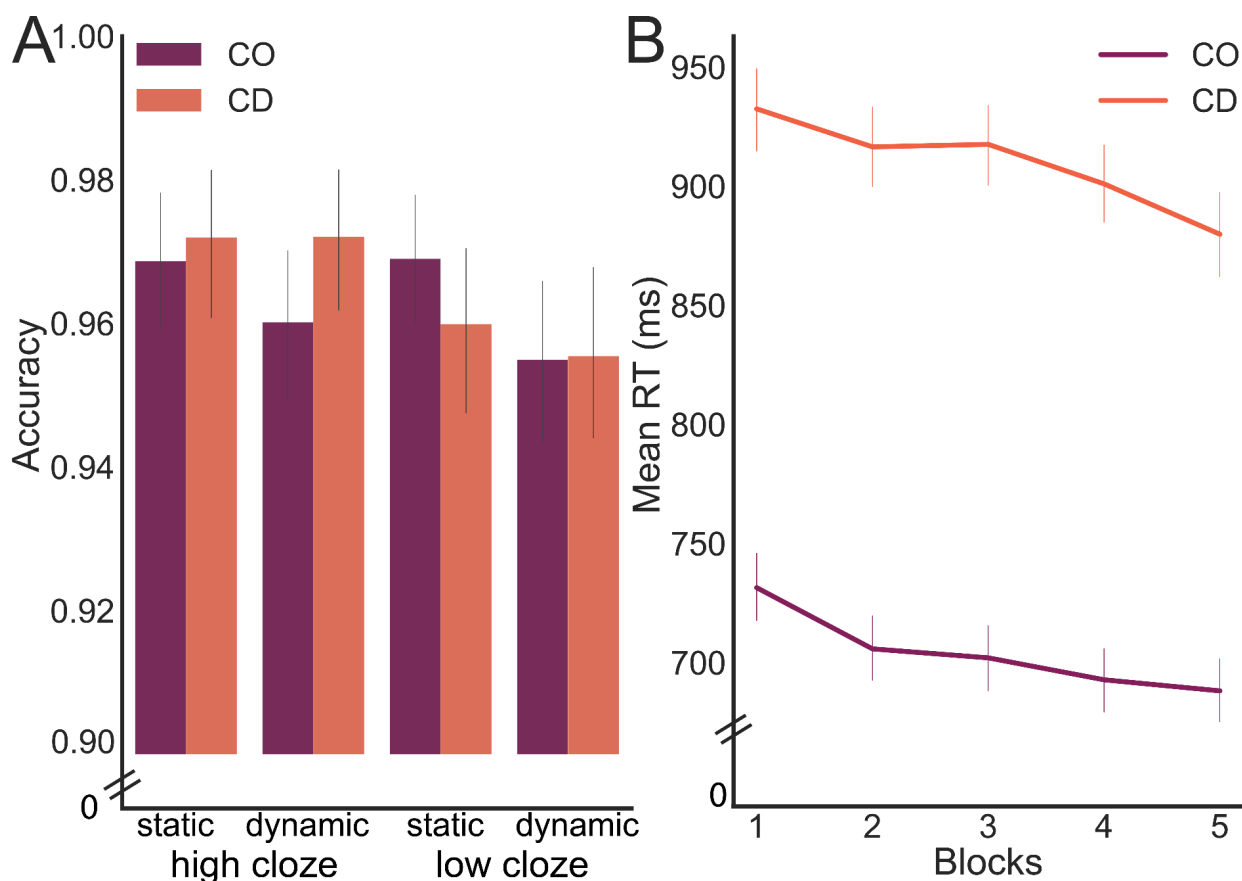


Fig 1. Behavioral performance summary for the cerebellar degeneration (CD) and control (CO) groups. A) Mean accuracy averaged across blocks is close to ceiling performance for both groups. B) Mean reaction time on correct trials for meaningful sentences as a function of test block (64 trials/block). The CD group responded slower than the CO group, with both groups showing a reduction in RT over the experimental session.

Our main interest centers on how the semantic judgments were influenced by the semantic content of the base phrases. Specifically, we asked how the degree of prediction and dynamic simulation influenced participants' response times on the verification task. We first focused on the meaningful sentences given that the time to process the final word should be faster when that word is predicted by the context. There was a main effect of Prediction ($F_{1,120} = 12.55$, $p < 0.001$, $\eta^2 = 0.09$), with faster response times on trials in which the target word had high cloze probability compared to trials in which the target word had low cloze probability (Fig 2A,B; Fig S1). Importantly, the Group X Prediction interaction was not significant ($F_{1,120} = 0.008$, $p = 0.93$, $\eta^2 < .001$): Participants in the CD group showed a similar benefit as the CO participants when the base phrase enabled the anticipation of the target word. Thus, this analysis fails to support the prediction derived from the internal model hypothesis, namely that the difference between high and low cloze conditions would be attenuated in the CD group.

Turning to the CoRT hypothesis, we next turned to the comparison of verification times when the base phrase described a dynamic or static context. As noted in the Introduction, we had no *a priori* reason to suppose that verification times for the control group would be sensitive to the relative dynamics; our focus was on whether the CD group would be selectively affected on the dynamic sentences. For both groups, the results indicate that RTs did not differ for sentences that described a dynamic event compared to sentences that described a static event ($F_{1,120} = 0.02$, $p = 0.89$, $\eta^2 < .001$, Fig 2C,D; Fig S1). Importantly, this factor did not interact with Group ($F_{1,120} = 0.07$, $p = 0.79$, $\eta^2 < .001$). Thus, the results fail to support the hypothesis that the integrity of the cerebellum may be disproportionately important when the semantic content engages a dynamic mental transformation. The 3-way Group x Prediction x CoRT interaction was also not significant ($F_{1,120} = 0.04$, $p = 0.85$, $\eta^2 < .001$)

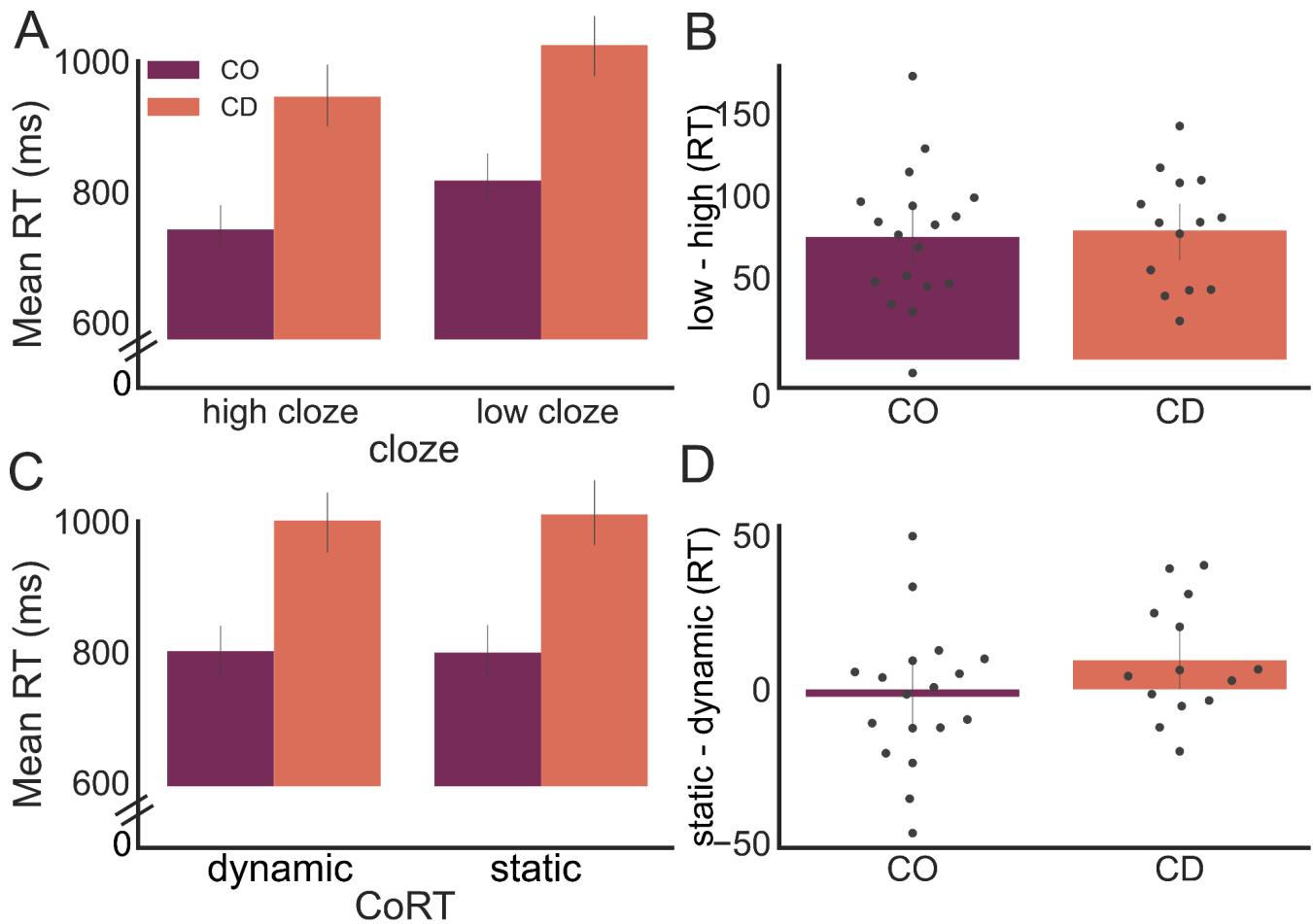


Fig 2. No differences between the CD and CO groups in their ability to use predictive semantic information. A) Mean RTs to target words as a function of cloze probability. B) Benefit on verification task to target words that are highly predictable (Low Cloze – High Cloze). Both the CD and CO groups were faster when the target word had high predictability, and the benefit was similar for the two groups. C) Mean RTs to target words as a function of the dynamics conveyed by the base phrase. D) There was no difference in RT for the dynamic and static sentences, a null effect that was similar for both groups. RTs in all graphs are limited to trials in which the participants correctly verified that the sentence was meaningful.

We also examined the meaningless sentences to ask how the two groups responded to violations of semantic predictions. To address this question, we used a 3-way ANOVA, with the factors Group, Prediction, and Trial Type (meaningful vs meaningless). Response times were longer to target words that rendered the sentence meaningless compared to target words that resulted in a meaningful sentence (Fig 3, $F_{1,120} = 12.85$, $p < .01$, $\eta^2 = .097$). There was no main effect of Prediction for meaningless trials ($F_{1,60}$

= .39, $p = .54$, $\eta^2 < .01$) and Trial Type did not interact with Group ($F_{1,120} = .015$, $p = .91$, $\eta^2 < .001$) or Prediction ($F_{1,120} = 1.95$, $p = .17$, $\eta^2 < .001$). Thus, while prediction violations slowed response times, the results failed to show that this effect was sensitive to the degree of prediction conferred by the base phrase and, importantly, failed to identify any difference in how the CD and control groups respond to violations of semantic expectancies. (We note that although the target words were selected to yield semantic predictions, a few of the violations were syntactic rather than semantic. However, there were not enough syntactic violations to do a separate analysis of these sentences.)

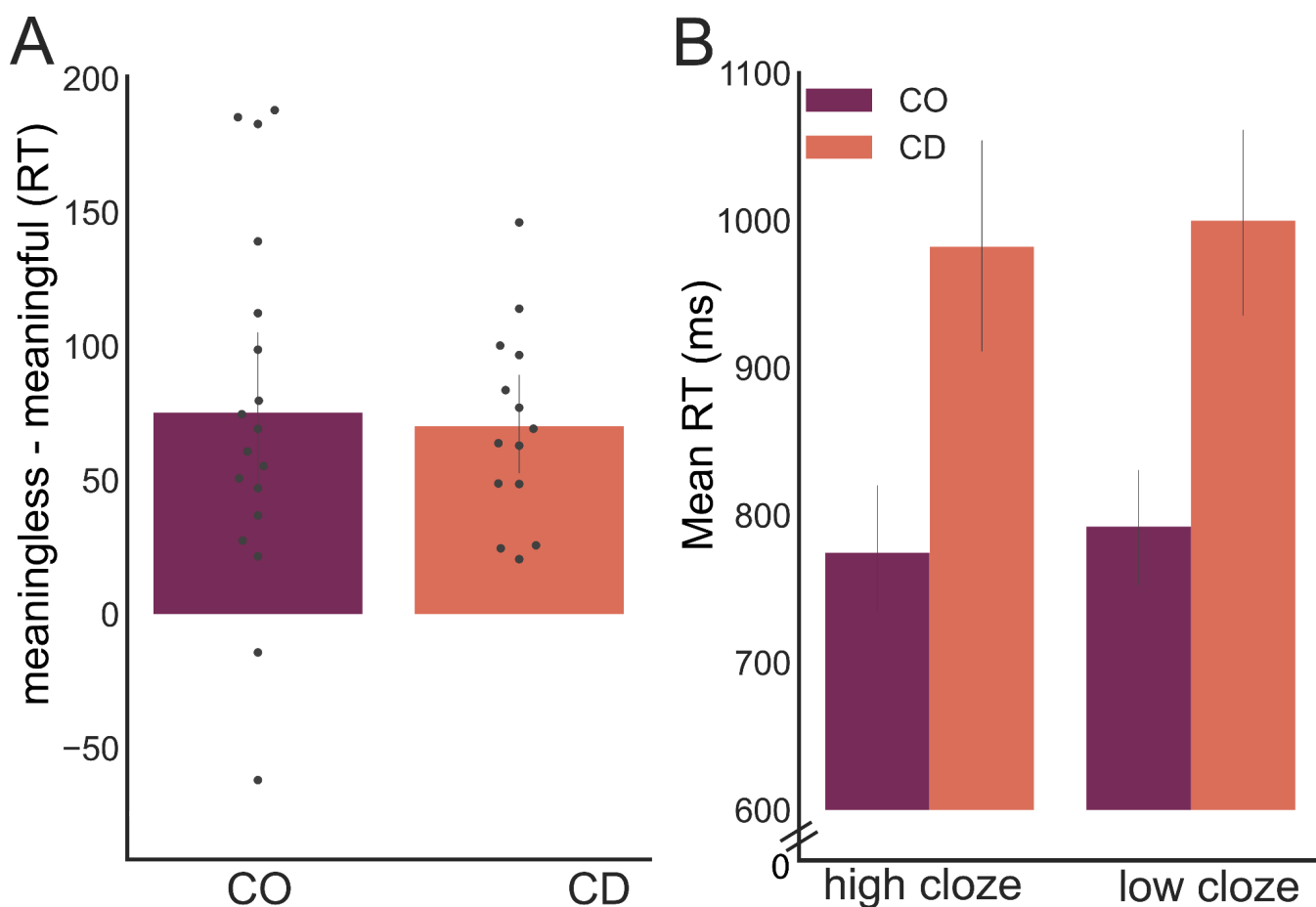


Fig 3. No differences between the CD and CO groups in evaluating sentences that entail prediction violations (i.e., meaningless trials). *A) RT difference between meaningless and meaningful sentences. Participants were slower when the target word did not fit with the context, and this effect was similar for the CO and CD groups. B) When the target word did not fit with the context, mean RT was numerically (but not statistically) faster when the context allowed for strong expectation (target word would have been high cloze). The magnitude of this effect was similar for the two groups.*

Discussion

To determine whether the cerebellum plays an important role in semantic processing, we tested patients with cerebellar degeneration (CD) and age- and education-matched healthy controls (CO) on a sentence verification task. Specifically, we were interested in evaluating two hypotheses concerning how the cerebellum might contribute to semantic processing. One hypothesis was based on the idea that the cerebellum forms internal models. Extending this idea in the language domain, we hypothesized that the integrity of the cerebellum would be required to generate predictions based on semantic content. The second hypothesis is based on the idea that the cerebellum is essential for mental operations that entail a continuous representational transformation (CoRT). When applying this constraint to language, we hypothesized that the integrity of the cerebellum would be essential for linguistic predictions that require a dynamic, rather than static, simulation of semantic content. The results failed to provide support for either hypothesis.

We parametrically manipulated semantic predictability (high vs low cloze probability), and the degree to which the semantic content suggested simulation (dynamic vs. static). We observed a main effect of prediction such that participants were faster to respond to sentences in which the target word was strongly predictable from the base phrase compared to sentences in which the target word was weakly predictable. This result corroborates previous literature that has reported an influence of cloze probability on reaction time (Staub et al. 2015). Despite this evidence that the task was sensitive to the operation of predictive mechanisms, the individuals with cerebellar degeneration showed response time advantage for the high cloze words. We also did not find support for the more constrained CoRT hypothesis, as we did not observe a selective cost for the CD group in response times to the sentences that might be expected to generate dynamic simulations. Moreover, the results provide indirect evidence against another task-independent hypothesis of cerebellar function, error-based processing (Fiez, 1992). Participants were slower to respond to the meaningless sentences compared to meaningful sentences (D’Mello et al. 2017). Taking this as a measure of sensitivity to prediction violations, the CD and control groups showed a similar increase in response time in judging a sentence as meaningless.

Using similar tasks, neuroimaging (Moberget et al. 2014; D’Mello et al. 2017; Lesage et al. 2017) and neurostimulation studies (D’Mello et al. 2017; Lesage et al. 2012) have produced evidence to suggest a cerebellar role in semantic prediction, with the emphasis in those studies on the internal model

hypothesis. For example, compared to scrambled word strings, predictive sentences produce an increase in the BOLD response in a circumscribed region of posterolateral right cerebellum, and violations of these predictions produce a strong response that is roughly symmetric for right and left posterolateral cerebellum (Moberget et al. 2014). These results stand in contrast to the null results of the present study, one in which we set out to provide a more direct test of a general and constrained variant of the internal model hypothesis through our parametric manipulations.

Taken together, we find an unsatisfying misalignment of neuroimaging, neurostimulation and neuropsychology findings in terms of how the cerebellum contributes to semantic processing. It is no doubt foolhardy to wish for a perfect marriage; different methods have their strengths and weaknesses. With respect to the cerebellum and cognition, much of the impetus has come from the neuroimaging literature, a method that is quite powerful in identifying areas that show activity correlated with tasks and/or processes. Neuropsychology offers a powerful tool to test functional hypotheses, although neural plasticity places some limitations in interpretation, especially for null results. It is also important to keep in mind one important limitation with cerebellar neuroimaging: The BOLD signal appears to be a function of activity at the mossy fiber terminals (Mathiesen et al. 2000) rather than activity reflective of intra-cerebellar processing (e.g., Purkinje cell firing). Thus, cerebellar neuroimaging is providing a window concerning the inputs to the cerebellum, but mute in terms of how that information is transformed by the cerebellum.

Caution is, of course, warranted when interpreting null results. We can consider a few possible explanations for why we failed to observe support for either hypothesis. First, we do note that there were performance differences between the CD and control groups, namely that the CD group was slower overall. This is to be expected given their ataxia. In previous studies involving manual responses for choice RT tasks, we have observed an increase in reaction time of between 50 and 150 ms (Helmuth et al. 1997; Breska & Ivry, 2018). In the present study, the increase was 220 ms. It is possible that the larger difference in the present study is motoric in nature, related to the fact the testing was performed on-line with a non-optimal response device (computer keyboard) for individuals with ataxia. On the other hand, it is possible that the increase in response time does reflect an impairment in sentence comprehension or some more generic aspect of cognitive processing, and that our experimental manipulations are insensitive to this impairment.

Second, our patient sample is quite heterogeneous, composed of individuals with different etiologies. It is possible that, when treated as a group, the pathology does not consistently impact critical regions within the cerebellum that are essential for semantic predictions (e.g., right posterolateral cerebellum). However, it is unlikely that the null results are attributable to divergent patterns of degeneration as the literature indicates that pathology in posterolateral regions of the cerebellum are generally associated with the etiologies of our sample (Lukas et al. 2006; Hernandez-Castillo et al. 2018). Nonetheless, it would be useful to recruit a larger sample size in future work to allow for voxelwise morphometry to relate patterns of pathology and behavior (Hernandez-Castillo et al. 2018; Kansal et al. 2017). We were not able to do any analysis along these lines in the present study: Not only is our sample small, but we do not have scans for most of the participants. Another approach would be to recruit individuals with unilateral lesions encompassing posterolateral regions, and compare the effect of right and left cerebellar lesions.

Another approach might be to focus on individuals with cerebellar degeneration who show language impairments on standard neuropsychological tests of language (e.g., range of semantic fluency tasks, word completion, Boston Naming Test (Kaplan et al. 1983)). This information would allow us to ask if performance on our experimental task of semantic prediction relates to profiles of language competence as assessed by these instruments. While our screening test for cognitive impairment (MoCA) included a semantic fluency task, it is likely neither comprehensive nor sensitive enough to detect subtle language deficits. We note that the patient group in the current study presented with relatively mild cognitive symptoms (as reported by MoCA). It is possible that individuals with milder impairment are more likely to participate in online studies.

Third, our experimental design may not have been optimal for testing the prediction and/or CoRT hypotheses. For example, we imposed a 500 ms delay prior to the presentation of the target word, with this interval serving as a cue that the next word would be the target. However, the delay period may have masked the cost of an impaired prediction process, one that operates more slowly in the CD group compared to the CO group. That is, the delay period could have provided a “slack” window that provided sufficient time for the CD group to generate a robust semantic prediction (Pashler, 1994). This hypothesis can be addressed by repeating the experiment without the delay period, perhaps using a

change in color or font to designate the target word. In terms of the CoRT hypothesis, sentence processing in general may entail the dynamic manipulation of the constituent words and phrases, independent of the semantic content of the sentences. By this view, a CoRT impairment might impact both our dynamic and static sentences. Tasks involving word pair judgments would present another means to compare semantic comprehension without imposing the more temporally-extended demands of sentence processing. For example, would the CD group show weaker priming for dynamic word pairs (ball-throw) compared to static word pairs (sunset-red).

Fourth, drawing on the motor control literature, we have conceptualized prediction as a dynamic process, one in which an input is fed into an internal model to generate an expectancy of the expected sensory consequences (i.e., the target word). It is possible that semantic expectancies do not entail simulation of this form, but rather emerge from memory retrieval (Firth, 1953; Mikolov, 2013; Smith & Levy, 2013). That is, after hearing the phrase, “The boss refused to give him a ---“, anticipating the word “raise” may not require running a simulation; rather, “raise” may be primed because it tends to co-occur (sadly) with the words “boss” and “refused.” Indeed, a memory-based account might be especially appropriate for high cloze sentences given that they tend to be more formulaic than low cloze sentences (Conklin & Schmitt, 2012). In general, cerebellar degeneration does not impact memory retrieval (Appollonio et al. 1993; McDougle et al. 2022). The similar sensitivity of the CD and control groups to cloze probability in the current study may be another manifestation of spared memory and memory retrieval in cerebellar degeneration.

Conclusions

Dating back to the earliest neuroimaging studies, language studies have figured prominently in the quest to understand the role of the cerebellum in cognition. Many of the studies have involved tasks that tax semantic retrieval, with the results showing consistent activation in right posterolateral cerebellum when people generate semantic associates (Petersen et al. 1989), manipulate information in verbal working memory (Marvel & Desmond, 2010), or verify meaningfulness (Moberget et al. 2014; D’Mello et al. 2017). Given that cerebellar pathology is not associated with profound aphasia, functional hypotheses have focused on how the cerebellum facilitates linguistic processing. Prediction has been central to most of these hypotheses given that fluent communication requires feedforward processing, defined within the domain of language comprehension as the ability to anticipate linguistic intent that arises during a

conversation with another individual or reading a text. The internal model and CoRT hypotheses provide two ways in which the cerebellum might support prediction in the domain of language, and indeed, support cognition more generally (Diedrichsen et al., 2018; McDougle et al., 2022). However, we failed to obtain support for either hypothesis when put to what we see as rigorous tests. We have outlined limitations with our study, and certainly see a need for future studies that provide further tests of the internal model and CoRT hypotheses, as well as other dimensions of semantic processing. We also recognize that prediction, at least as conceptualized here, may not be the appropriate kernel for understanding how the cerebellum supports language.

Data Availability

All of the sentences used across experimental task blocks (including cloze and CoRT ratings) are available to download on open science framework at https://osf.io/aegkm/?view_only=None. The code used to analyze these data are available to download on Github at https://github.com/maedbhk/NOL_cerebellum.

References

1. Ackermann, H., & Hertrich, I. (1994). Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing. *Folia Phoniatrica et Logopaedica*, *46*(2), 70-78.
2. Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain and language*, *60*(2), 323-331.
3. Alexander, M. P., Gillingham, S., Schweizer, T., & Stuss, D. T. (2012). Cognitive impairments due to focal cerebellar injuries in adults. *Cortex*, *48*(8), 980-990.
4. Andreasen, N. C., Paradiso, S., & O'Leary, D. S. (1998). "Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in cortical-subcortical-cerebellar circuitry?. *Schizophrenia Bulletin*, *24*(2), 203-218.
5. Appollonio, I. M., Grafman, J., Schwartz, V., Massaquoi, S., & Hallett, M. (1993). Memory in patients with cerebellar degeneration. *Neurology*, *43*(8), 1536-1536.
6. Apps, R. & Hawkes, R. (2009). Cerebellar cortical organization: A one-map hypothesis. *Nature Reviews Neuroscience*, *10*(9), 670-681.

7. Argyropoulos, G. P. (2016). The cerebellum, internal models and prediction in ‘non-motor’ aspects of language: a critical review. *Brain and language*, *161*, 4-17.
8. Block, C. K. & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*(3), 665–670.
9. Breska, A., & Ivry, R. B. (2018). Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson’s disease. *Proceedings of the National Academy of Sciences*, *115*(48), 12283-12288.
10. Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45.
11. Dean, P., & Porrill, J. (2016). The importance of Marr’s three levels of analysis for understanding cerebellar function. *Computational Theories and their Implementation in the Brain: The legacy of David Marr*, 79.
12. Desmond, J. E., & Fiez, J. A. (1998). Neuroimaging studies of the cerebellum: language, learning and memory. *Trends in cognitive sciences*, *2*(9), 355-362.
13. Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M., & Ivry, R. B. (2019). Universal transform or multiple functionality? Understanding the contribution of the human cerebellum across task domains. *Neuron*, *102*(5), 918-928.
14. D’Mello, A. M., Turkeltaub, P. E., & Stoodley, C. J. (2017). Cerebellar tDCS modulates neural circuits during semantic prediction: a combined tDCS-fMRI study. *Journal of Neuroscience*, *37*(6), 1604-1613.
15. Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?. *Neural Networks*, *12*(7-8), 961-974.
16. Eccles J. C., Ito M., Szentágothai J. (1967). *The cerebellum as a neuronal machine*. Oxford, England: Springer-Verlag.
Fiez, J. A., Petersen, S. E., Cheney, M. K., & Raichle, M. E. (1992). Impaired non-motor learning and error detection associated with cerebellar damage: A single case study. *Brain*, *115*(1), 155-178.
17. Fiez, J. A., Raichle, M. E., Balota, D. A., Tallal, P., & Petersen, S. E. (1996). PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cerebral cortex*, *6*(1), 1-10.

18. Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.
19. Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13(7), 293–301.
20. Gasparini, M., Piero, V. D., Ciccarelli, O., Cacioppo, M. M., Pantano, P., & Lenzi, G. L. (1999). Linguistic impairment after right cerebellar stroke: a case report. *European Journal of Neurology*, 6(3), 353-356.
21. Helmuth, L. L., Ivry, R. B., & Shimizu, N. (1997). Preserved performance by cerebellar patients on tests of word generation, discrimination learning, and attention. *Learning & Memory*, 3(6), 456-474.
22. Hernandez-Castillo, C. R., King, M., Diedrichsen, J., & Fernandez-Ruiz, J. (2018). Unique degeneration signatures in the cerebellar cortex for spinocerebellar ataxias 2, 3, and 7. *NeuroImage: Clinical*, 20, 931-938.
23. Holmes, G. (1917). The symptoms of acute cerebellar injuries due to gunshot injuries. *Brain*, 40(4), 461-535.
24. Holmes, G. (1939). The cerebellum of man. *Brain*, 62(1), 1-30.
25. Ito, M. (2008). Controls of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, 9(4), 304-313.
26. Kansal, K., Yang, Z., Fishman, A. M., Sair, H. I., Ying, S. H., Jedynak, B. M., ... & Onyike, C. U. (2017). Structural cerebellar correlates of cognitive and motor dysfunctions in cerebellar degeneration. *Brain*, 140(3), 707-720.
27. Keele, S. W. & Ivry, R. (1990). Does the Cerebellum Provide a Common Computation for Diverse Tasks? A Timing Hypothesis. *Annals of the New York Academy of Sciences* 608(1), 179–211.
28. Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163
29. Ivry, R. B., & Fiez, J. A. (2000). Cerebellar contributions to cognition and imagery. *The new cognitive neurosciences*, 2, 999-1011.
30. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. (2019). Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature Neuroscience*, 22(8), 1371–1378.

31. Leiner, H. C., Leiner, A. L. & Dow, R. S. (1986). Does the Cerebellum Contribute to Mental Skills? *Behavioral Neuroscience*, 100(44), 443–454.
32. Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325-343.
33. Lesage, E., Hansen, P. C., & Miall, R. C. (2017). Right lateral cerebellum represents linguistic predictability. *Journal of Neuroscience*, 37(26), 6231-6241.
34. Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar rTMS disrupts predictive language processing. *Current Biology*, 22(18), R794-R795.
35. Lesage, E., Nailer, E. L. & Miall, R. C. (2016). Cerebellar BOLD signal during the acquisition of a new lexicon predicts its early consolidation. *Brain and Language*, 161, 33–44.
36. Lukas, C., Schöls, L., Bellenberg, B., Rüb, U., Przuntek, H., Schmid, G., ... & Suchan, B. (2006). Dissociation of grey and white matter reduction in spinocerebellar ataxia type 3 and 6: a voxel-based morphometry study. *Neuroscience letters*, 408(3), 230-235.
37. Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.
38. Marvel, C. L., & Desmond, J. E. (2010). Functional topography of the cerebellum in verbal working memory. *Neuropsychology review*, 20(3), 271-279.
39. Mathiesen, C., Caesar, K. & Lauritzen, M. Temporal coupling between neuronal activity and blood flow in rat cerebellar cortex as indicated by field potential analysis. (2000).
40. McDougale, D.S., Tsay, J.S., Pitt, B., King, M., Saban, W., Taylor, J.A., & Ivry, R.B. (2022). Continuous manipulation of mental representations is compromised in cerebellar degeneration, *Brain*.
41. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
42. Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B. & Endestad, T. (2014). Generalized role for the cerebellum in encoding internal models: Evidence from semantic processing. *The Journal of Neuroscience*, 34(8), 2871–2878.
43. Moberget, T., Hilland, E., Andersson, S., Lundar, T., Due-Tønnessen, B. J., Heldal, A., Ivry, R.B., & Endestad, T. (2016). Patients with focal cerebellar lesions show reduced auditory cortex activation during silent reading. *Brain and language*, 161, 18-27.

44. Moberget, T. & Ivry, R. B. (2016). Cerebellar contributions to motor control and language comprehension: searching for common computational principles. *Annals of the New York Academy of Sciences*, 1369(1), 154-171.
45. Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, 116(2), 220.
46. Peelle, J. E., Miller, R. L., Rogers, C. S., Spehar, S., Sommers, M. S., Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. *Behavior Research Methods*.
47. Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of Cognitive Neuroscience*, 1(2), 153-170.
48. Pickering, M. J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110.
49. Poort, E. D., & Rodd, J. M. (2019). Towards a distributed connectionist account of cognates and interlingual homographs: Evidence from semantic relatedness tasks. *PeerJ*, 7.
50. Rabe, K., Livne, O., Gizewski, E. R., Aurich, V., Beck, A., Timmann, D., & Donchin, O. (2009). Adaptation to visuomotor rotation and force field perturbation is correlated to different brain areas in patients with cerebellar degeneration. *Journal of neurophysiology*, 101(4), 1961-1971.
51. Ravizza, S. M., McCormick, C. A., Schlerf, J. E., Justus, T., Ivry, R. B., & Fiez, J. A. (2006). Cerebellar damage produces selective deficits in verbal working memory. *Brain*, 129(2), 306-320.
52. Richter, S., Kaiser, O., Hein-Kropp, C., Dimitrova, A., Gizewski, E., Beck, A., Aurich, V., Ziegler, W., & Timmann, D. (2004). Preserved verb generation in patients with cerebellar atrophy. *Neuropsychologia*, 42(9), 1235-1246.
53. Richter, S., Gerwig, M., Aslan, B., Wilhelm, H., Schoch, B., Dimitrova, A., Gizewski, E. R., Zielger, W., Karthath, H. O., & Timmann, D. (2007). Cognitive functions in patients with MR-defined chronic focal cerebellar lesions. *Journal of Neurology*, 254(9), 1193-1203.
54. Riva, D. (1998). Language deficits in a child with omolateral (left) temporo-basal and cerebellar lesions. *Neuropsychologia*, 36(1), 71-75.
55. Schmahmann, J. D. (1996). From movement to thought: anatomic substrates of the cerebellar contribution to cognitive processing. *Human Brain Mapping*, 4(3), 174-198.

56. Schmahmann, J. D. (1998). Dysmetria of thought: clinical consequences of cerebellar dysfunction on cognition and affect. *Trends in Cognitive Sciences*, 2(9), 362-371.
57. Schmahmann, J. D. & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain*, 121(4), 561–579.
58. Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review of Neuroscience*, 33, 89-108.
59. Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701-703.
60. Smith NJ, Levy R. Cognition. 2013 Sep;128(3):302-19.
61. Sokolov, A. A., Miall, R. C., & Ivry, R. B. (2017). The cerebellum: adaptive prediction for movement and cognition. *Trends in cognitive sciences*, 21(5), 313-332.
62. Spencer, R. M., Zelaznik, H. N., Diedrichsen, J., & Ivry, R. B. (2003). Disrupted timing of discontinuous but not continuous movements by cerebellar lesions. *Science*, 300(5624), 1437-1439.
63. Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, 82, 1-17.
64. Stoodley, C. J., & Schmahmann, J. D. (2009). The cerebellum and language: evidence from patients with cerebellar degeneration. *Brain and language*, 110(3), 149-153.
65. Stoodley, C. J., Valera, E. M., & Schmahmann, J. D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *Neuroimage*, 59(2), 1560-1570.
66. Day, B. L., Thompson, P. D., Harding, A. E., & Marsden, C. D. (1998). Influence of vision on upper limb reaching movements in patients with cerebellar ataxia. *Brain: a journal of neurology*, 121(2), 357-372.
67. Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880-1882.
68. Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9), 338-347.