The right lateral cerebellum represents linguistic predictability.

Elise Lesage\textsuperscript{1,2}, Peter C Hansen\textsuperscript{1}, and R Chris Miall\textsuperscript{1}

Author names and affiliation:

1: School of Psychology, University of Birmingham, UK

2: Currently at: NIDA IRP, National Institutes of Health, Baltimore US

Key words: cerebellum, language, comprehension, non-motor, prediction, cloze probability, fMRI, prediction error, phonological working memory

Corresponding author with complete address, including an email address and postal code

Elise Lesage

Postal address:

Neuroimaging Research Branch, NIDA/NIH

Baltimore, MD, 21202

Email: elise.r.d.lesage@gmail.com
Conflict of Interest: none to declare

Acknowledgements: The work was funded by EU-FP7-ITN C7 and by Wellcome Trust grant WT087554.

The authors would like to thank Dr. Jessica Bernard for proving the Crus I and Crus II connectivity maps.
Abstract

Mounting evidence indicates that posterolateral portions of the cerebellum (right Crus I/II) contribute to language processing, but the nature of this role remains unclear. Based on a well-supported theory of cerebellar motor function, which ascribes to the cerebellum a role in short-term prediction through internal modeling, we hypothesize that right cerebellar Crus I/II supports prediction of upcoming sentence content. We tested this hypothesis using event-related fMRI in human subjects by manipulating the predictability of written sentences. Our design controlled for motor planning and execution, as well as for linguistic features and working memory load; it also allowed separation of the prediction interval from the presentation of the final sentence item. In addition, three further fMRI tasks captured semantic, phonological and orthographic processing, to shed light on the nature of the information processed. As hypothesized, activity in right posterolateral cerebellum correlated with the predictability of the upcoming target word. This cerebellar region also responded to prediction error during the outcome of the trial. Further, this region was engaged in phonological, but not semantic or orthographic processing. This is the first imaging study to demonstrate a right cerebellar contribution in language comprehension independently from motor, cognitive and linguistic confounds. These results complement our work using other methodologies showing cerebellar engagement in linguistic prediction, and suggest that internal modeling of phonological representations aids language production and comprehension.
Significance statement

The cerebellum is traditionally seen as a motor structure that allows for smooth movement by predicting upcoming signals. However, the cerebellum is also consistently implicated in non-motor functions such as language and working memory. Using fMRI, we identify a cerebellar area that is active when words are predicted and when these predictions are violated. This area is active in a separate task that requires phonological processing, but not in tasks that require semantic or visuospatial processing. Our results support the idea of prediction as a unifying cerebellar function in motor and non-motor domains. We provide new insights by linking the cerebellar role in prediction to its role in verbal working memory, suggesting that these predictions involve phonological processing.
**Introduction**

The cerebellar role in language and cognition has become increasingly apparent over recent decades (Strick et al., 2009). Patient and functional imaging data show that cerebellar regions contributing to language and cognition are largely confined to the posterolateral cerebellum (hemispheric portions of Lobule VII, consisting of Crus I and Crus II). These regions are reciprocally connected with supramodal neocortical areas, as demonstrated using tracer studies in non-human primates (Kelly and Strick, 2003) and by functional connectivity MRI in humans (Buckner et al., 2011; Bernard et al., 2012). A wealth of neuroimaging studies report right posterolateral cerebellar activation in studies that probe language (Stoodley and Schmahmann, 2010; Price, 2012) and working memory processes (Desmond et al., 1997; Hayter et al., 2007; Stoodley and Schmahmann, 2009; Keren-Happuch et al., 2012).

However, the functional contribution of the cerebellum in language remains unclear. In motor control, the cerebellum is thought to acquire and store internal models of the motor system. These internal models predict upcoming reafferent sensory input, and these continuous short-term predictions allow for fluent movements and efficient error correction (Miall et al, 1993, Wolpert and Miall, 1996; Miall, 1998). Based on the homogeneous cerebellar cytoarchitecture, several authors have argued that the cerebellar role in non-motor functions is like that in motor control, performing similar operations on more abstract inputs.
(Bloedel, 1992; Ramnani, 2006; Ito, 2008). Thus, extrapolating from the internal model motor theory of
the cerebellum, the posterolateral areas of the cerebellum might support short-term prediction of future
linguistic stimuli.

A testable hypothesis can be derived from this proposal: the cerebellum, specifically right Crus I/II, should
be differentially engaged when processing highly predictable versus unpredictable language. Consistent
with this notion, online prediction of upcoming sentence content is slowed after perturbation of the right
cerebellum with transcranial magnetic stimulation (TMS; Lesage et al., 2012) and modulated by electrical
stimulation (tDCS; Miall et al., 2016; D’Mello et al., 2017). In addition, fMRI studies have reported right
cerebellar recruitment in conditions where linguistic prediction is possible (Desmond et al., 1998;
Moberget et al., 2014). However, it has been difficult to manipulate linguistic prediction without also
introducing differences in speech production processes, linguistic properties of the stimulus, task difficulty
(working memory load), or outcome evaluation (prediction error); each of these processes have been
shown to recruit the posterior cerebellum (Petersen et al., 1989; Floyer-Lea and Matthews, 2004;
Fedorenko et al., 2010; Stoodley et al., 2012; Grimaldi et al., 2014; Argyropoulos, 2015; Moberget and
Ivry, 2016). To date, no fMRI study has been able to capture cerebellar responses to linguistic prediction
during comprehension while controlling for these confounds.
Here, we manipulate the predictability of sentences in an event-related fMRI design, and test whether the haemodynamic response in right Crus I/II covaries with predictability. Critically, the time at which a prediction is made was isolated from the outcome of the sentence and from the contextual information that allows a prediction to be made. In addition, we explored whether the cerebellar roles in working memory and linguistic prediction could be reconciled; e.g. perhaps linguistic prediction requires short-term storage of semantic, phonological or orthographic representations. Thus, we further assessed whether cerebellar regions identified in the predictive task were engaged in three additional fMRI tasks which capture semantic, phonological and orthographic (visuospatial) working memory.

Materials and methods

Participants

Eighteen right-handed volunteers (4 male, average age 21 years, range 18-27 years) participated in two fMRI sessions. One male subject was excluded from the second session and from all data analysis due to severe signal dropout in the lateral cerebellum. All participants were native English speakers; none were fluent in any other language. Participants were remunerated for their time. Written informed consent was obtained for each participant. This study was approved by the local ethics committee at the University of
Birmingham and was carried out in accordance with the guidelines set out in the Declaration of Helsinki (1964).

**Prediction task**

Participants silently read visually presented sentences with varying degrees of predictability, and pressed an MR-compatible response button to indicate the plausibility of the sentence. Participants were not informed that the predictability of sentences was relevant, and were merely instructed to read the words presented on the screen and judge whether the outcome of the final item was likely given the context.

The task consisted of 78 trials, each presenting a unique item (context sentence + stem of a second sentence). Thirty three items were taken from a study by Fitzsimmons and Drieghe (2013) and altered to better suit this fMRI design; 45 items were newly constructed. A behavioral pilot experiment in an independent sample of 43 participants had determined the items’ predictability (cloze probability). Cloze probability can be defined as the probability that a sentence will be completed with a given target word (e.g. a cloze probability of 0.90 indicates that 90% of participants will complete the item with the same target word). Cloze probability was used as a continuous parametric modulator in behavioral and fMRI analysis. We also categorized items as neutral (cloze probabilities between 0 and 0.40; 27 items), semi-predictable (cloze probabilities between 0.40 and 0.70; 25 items) and predictable (cloze probabilities
between 0.70 and 1.00; 26 items). These discrete levels of predictability were used for easier visualization of the results; all analyses were conducted with cloze probability as a continuous variable.

Three temporal events per trial were independently modelled to allow separate estimation of the BOLD response to these events (Figure 1). The first was the presentation of a context sentence (CONTEXT), which appeared on the screen for three seconds (e.g. "Sonja wanted to avoid a sunburn in this hot weather."). Context sentences were controlled for the number of syllables and words. The second event was the presentation of the stem of a second sentence (STEM; e.g. "She had brought some …"). The stem was displayed in 4 parts (consisting of one or two whole words), each displayed for 250ms in the center of the screen to avoid eye movements. The stem did not contain the last word of the sentence, and it is inferred that the participant would produce a semantic prediction (e.g. "sunscreen") in a highly predictive item. Thus, prediction and predictability are measured at the time of the STEM event, before the final word. The third event in the trial was the presentation of the final word of the sentence (OUTCOME), which was either likely (50% of trials) or unlikely (50% of trials) given the context.

Participants then made a response on a MR-compatible response box to indicate plausibility. Importantly, whether the outcome was likely or unlikely was independent from how predictable the item was. Highly predictable and unpredictable items could be paired with a likely or unlikely outcome. The STEM, and the
inferred prediction at its end, is the event of interest in this task. Items were constructed in pairs and triplets so that a similar sentence stem was used for different levels of predictability. The length and linguistic properties of the STEM were therefore well-controlled between conditions. Presentation of the outcome and the button-press response were modeled as a single event (OUTCOME, 1s), ensuring that prediction error, motor preparation or motor activity could not contribute to the haemodynamic response at the time of the STEM. Trials with erroneous responses were excluded from the fMRI analysis.

Uniformly distributed variable delays were introduced between context and stem (4.5 - 10.5 seconds), between stem and outcome (3 - 7.5 seconds), and between outcome and the context event of the following trial (4-10 seconds). This manipulation ensured that BOLD responses to one event were not contaminated with BOLD response to the previous stimulus (for another example of this technique, see Ramnani and Miall, 2004).

Localizer tasks for semantic, phonological and orthographic working memory

When reading a sentence (or a sentence stem), processes in addition to semantic prediction take place.

When reading words, one processes the meaning of these words (attention to semantics). When reading
words or pronounceable non-words, one processes phonological features of these words (attention to phonology). When looking at words or non-words, one recognizes and processes a visual stimulus with a certain spatial configuration (attention to orthography or visuospatial attention). In order to assess whether any cerebellar areas that differentially respond to predictive sentences were also preferentially engaged when orthographic, phonological or semantic properties were held in short-term store, participants also performed three epoch-related localizing tasks.

To maximize comparability between tasks and to have low level of working memory load, all three tasks were 1-back tasks and were contrasted with 0-back versions of the same task. The participants were required press a button on an MRI compatible response box if the current stimulus belonged to the same semantic category as the previous stimulus (semantic 1-back), if the current stimulus rhymed with the previous stimulus (phonological 1-back) or if the current stimulus was identical to the previous stimulus (orthographic 1-back). Similar tasks have previously been used to capture orthographic and phonological processing (Paulesu et al., 1993; Koyama et al., 2013). Three 0-back control conditions required the participants to respond when a known target stimulus appeared. The 0-back controls blocks were performed as separate runs from the 1-back blocks.

Semantic 1-back. For the semantic task, stimuli were 50 black-and-white line drawings. Participants were familiarized with the ten stimulus categories (cycles, birds, boats, dogs, fish, fruits, buildings, shoes, tools...
and furniture) and the five members of each category, as well as with the 0-back task target object, prior
to scanning (Figure 2A). In contrasting the 1-back with the 0-back condition, we control for visual
processing of the line drawings, and motor activity related to button presses. The requirements that
separate the 1-back condition from the 0-back condition are that in the 1-back condition, participants must
categorize each stimulus, keep this semantic category in short-term memory, and match it to the semantic
category of the subsequent stimulus. In the 0-back condition it is not necessary to process the meaning or
semantic category of the line drawing, merely to match it to a target image. We chose line drawings
instead of words to avoid automatic phonological processing; line drawings hold meaning but are non-
verbal. Nevertheless, we cannot exclude that participants formed a phonological code of the stimulus or
the semantic category.

Phonological 1-back. For the phonological task, stimuli were 5-letter words that were printed in the middle
of the screen (Figure 2C). Before the scanning session, participants were shown some example stimuli
for this task, which were not used in the scanning task. They were also shown the target stimulus for the
0-back task. The task was constructed such that a small minority of the rhyming pairs ended in the same
syllable. This task could therefore not be performed to an acceptable standard by using a visual search
strategy. The 1-back and 0-back conditions were controlled for reading requirements (each condition
required reading 5-letter words), demands on attention to meaning or semantics (both conditions likely
automatically elicited semantic processing but neither condition required it) and motor activity related to the button presses. Unlike the 0-back condition, the 1-back condition required participants to update and hold the phonetic form of each stimulus in short-term storage and match it to the phonetic form of the subsequent stimulus. The 0-back condition merely required the subject to hold in memory and respond to one target word (the word "press") throughout the run.

Orthographic 1-back. For the orthographic (visuospatial) task, stimuli consisted of a set of 10 five-letter Punjabi pseudo-words (Figure 2C). These stimuli had a similar configuration as written English words, but held no meaning and were not pronounceable for the participants. Participants were familiarized with all the visual stimuli, as well as with the target stimulus for the 0-back task prior to scanning. Both conditions were matched for low-level visual demands as well as motor activity related to button presses. As in the phonological task, the difference in requirement for the 1-back condition was the higher short-term memory load to retain the visuospatial configuration of each stimulus and compare it to the subsequent stimulus, whereas the 0-back condition required only one easy to identify stimulus to be retained throughout.

Each of the six runs (3 tasks, each as 1-back and 0-back) lasted 8 minutes and contained 15 epochs. Each epoch consisted of 10 stimuli and lasted 15 seconds. Stimuli were presented for 500ms, 1000ms
apart. Rest periods between blocks lasted 13 to 17 seconds. These rest periods (53% of the scan) were used as an implicit baseline in the analysis.

*MRI data acquisition*

Each participant underwent two fMRI scanning sessions on separate days. One session consisted of a prediction task, divided into 3 runs each lasting 10min 30s. A high-resolution structural image (T1 weighted image, FTE sequence, voxels 1x1x1mm) was also collected during this session. During a second fMRI session, participants performed three localizer tasks, designed to probe attention to semantic, phonological and orthographic features of visually presented stimuli. Localizer tasks were divided into an experimental run (1-back condition) and a control run (0-back condition), with each run lasting 8 mins. Runs were presented in the same order for each participant. All images were acquired on a 3T Philips Achieva scanner using a 32-channel head coil. (Functional: ascending EPI sequence, TR=3s, TE=32ms, 52 axial slices (no gap), voxels 3x3x3 mm, FOV 240x240, flip angle = 85°). Pulse oximetry and breathing measures were collected with a Philips-integrated physiological monitoring system.
Behavioral data were processed using custom-made MATLAB scripts (RRID:SCR_001622). Performance in the outcome phase of prediction task, as well as in the localizer tasks was analyzed in R (RRID:SCR_001905) using the packages afex and phia. For the prediction task, a generalized linear mixed model (random intercept, accuracy as binomial dependent variable) was carried out with Predictability (continuous cloze probability) and Outcome (levels: likely and unlikely) as independent variables. For the localizer tasks, a general linear mixed model (random intercept) was used with Task (levels: semantic, phonological, and orthographic) and Condition (levels: 1-back and 0-back). Significant interactions were followed up by post-hoc tests. Average performance was assessed in all conditions to ascertain that participants paid attention to the task, and to allow the exclusion of erroneous trials from the imaging analysis of the prediction task.

Statistical analysis of the fMRI data

Preprocessing. All analyses were carried out in SPM8 (RRID:SCR_007037). Prior to the first level analysis, raw images were realigned to correct for head motion, slice-time corrected, and co-registered to the anatomical image. First level analyses were carried out in subject-specific space. Contrast images
from the first level analysis were normalized to the SPM8 EPI template (whole-brain analysis) and
smoothed with an 8mm FWHM Gaussian smoothing kernel before entering group level analysis. To
facilitate later region-of-interest analyses, all EPI images were also normalized and smoothed. The BOLD
signal around the brainstem and cerebellum can be vulnerable to confounding physiological signals, but
this can be accounted for by regressing out heart rate and breathing signals in the GLM model (Schlerf et
al., 2012). The Physiological Log Extraction for Modeling (PhLEM) toolbox in SPM (Verstynen and
Deshpande, 2011) was used to convert heart rate and breathing traces into SPM regressors with a
CETROICOR method (Glover et al., 2000), resulting in eight regressors that were included as regressors
of no interest. Physiological measures from one participant during the control sessions were not available;
this person's data were excluded from the analysis of the control tasks.

First level analysis. Prediction task: For the linguistic prediction task, six events per block were modelled
at the first level. These were: context, context\text{mod} (a parametric modulator of the context by cloze
probability), stem, stem\text{mod} (a parametric modulator of the stem by cloze probability), outcome\text{likely}, and
outcome\text{unlikely}. The three blocks were concatenated, thus creating a single first level analysis per person
with 18 events of interest. A 19th regressor modeled all trials where an erroneous response was made to
ensure that differences in performance could not underlie differences in BOLD activation patterns. The six
contrasts of interest (the six events, averaged over the three blocks) were estimated against the implicit baseline.

Localizer task: For the localizer tasks, the task blocks were modeled against the implicit baseline in a single t-contrast for each of the six sessions.

In all tasks, eight regressors of no interest modeled physiological signals and a further six modeled head movement.

Group level analysis. Normalized first level contrast images for the prediction task were entered into a factorial design. First, the contrast \( t = \text{[stem]} \) (reading contrast), was estimated to assess which regions were engaged in the processing of written meaningful language, irrespective of predictability. Second, the predictability contrast, \( t = \text{[stem}_{\text{mod}} \) revealed areas where the BOLD signal was modulated according to the predictability of the upcoming sentence ending. A mask of the subjects' brains was created by averaging the normalized skull-stripped anatomical scans co-registered into a 2x2x2mm space (216,611 voxels, 1733cm³). A whole brain cluster-correction at a family-wise error rate (FWE) of 5% for this volume was calculated using the 3dclustsim algorithm (Cox, 1996). This procedure determined a voxel-level correction of \( p<0.001 \), with a minimum cluster size of 99 voxels (790mm³). In addition, we assessed whether cortical activations were in regions that are functionally connected with the cerebellar region of
interest. To this end, resting state connectivity maps with right Crus I and right Crus II (Bernard et al., 2012; maps provided by the authors) were summed and smoothed with a 4mm FWHM Gaussian smoothing kernel (Figure 4C). This resulting connectivity map was then overlaid with the activation map from the predictability contrast.

Region of interest (ROI) analyses on areas engaged in prediction

We conducted region-of-interest analyses to determine whether any cerebellar areas that are engaged in linguistic prediction also show increased activity when this prediction is violated (i.e. when the outcome is unlikely versus when it is likely; during a prediction error). Moreover, we further assess whether these cerebellar areas were engaged in semantic, phonological or orthographic processing in the three localizer tasks. Region of interest analyses were conduction using the marsbar toolbox in SPM8 (Brett et al., 2002). Regions of interest (ROIs) included cerebellar clusters that were modulated by predictability (predictability contrast), as well as cerebellar areas that were modulated by the presentation of written language (reading contrast). Given our a priori right cerebellar hypothesis, we planned to Bonferroni correct for the number of right cerebellar clusters that are identified in each contrast. In order to explore whether the activation patterns identified in the cerebellum were unique to this structure, or whether cerebral areas also showed the same patterns, we also plotted these parameter estimates of the
supratentorial clusters identified in the prediction contrast. These further ROI extractions are strictly
exploratory, and their results should not be interpreted. Masks of these areas were created by taking a
10mm sphere around the peak coordinate. First-level design matrices were accessed by marsbar to
extract the contrasts estimates for the regions of interest defined by the main analysis. This resulted in
one parameter estimate per participant per event per ROI.

Prediction error analysis. If linguistic internal models are present in the posterolateral cerebellum, one
might expect these regions to respond more strongly to the unlikely outcomes (prediction error) than to
the likely outcomes, analogous to the high activations seen when movement errors occur in motor tasks
(Imamizu et al., 2000; Miall et al., 2001). The first-level design matrix from the main prediction analysis
was used to extract parameter estimates for Outcome\(_{unlikely}\) and Outcome\(_{likely}\) events, which were then
compared with a paired t-test. An unlikely outcome does not mean that no prediction was made, it merely
means that the outcome violates expectations. A stronger response to unlikely versus likely outcomes
indicates that this region processes prediction errors. We hypothesized that those cerebellar areas that
are modulated by predictability also respond more strongly when a prediction is violated. A likely or
unlikely outcome was equally probable regardless of the item's predictability. This contrast was therefore
independent from the predictability contrast.
Localizer tasks: attention to semantics, phonology, and orthography. Given the recruitment of the posterolateral cerebellum in working memory tasks, we were interested to see whether those regions that are differentially engaged in linguistic prediction are also active in tasks that require short-term storage of semantic, phonological or orthographic stimulus features. Such functional overlap can provide us with insight into how the cerebellum contributes to language function, and how linguistic and working memory contributions may be reconciled. First level design matrices were created modeling the six conditions (1-back and 0-back conditions for the three localizing tasks) individually against the implicit baseline. Parameter estimates were extracted using marsbar and paired t-tests assessed whether the regions of interest showed a larger response to the 1-back condition than to the 0-back condition in the semantic, phonological and visual localizer. Data from the localizer tasks resulted from independent datasets (from the same participants). Circularity was therefore not a concern.

Results

Behavioral Results
Overall, participants performed well (average 86% correct, SEM = 2.5%, range 79-90%), indicating that all participants were attentive and able to judge whether a sentence ending was likely or unlikely in the context of the trial. The mixed-model ANOVA showed a significant effect for Predictability ($X^2_{1}=17.69$, $p<0.001$), Outcome ($X^2_{1}=15.48$, $p<0.001$) and their interaction ($X^2_{1}=8.24$, $p=0.004$). Post-hoc tests reveal that predictability did not affect performance on unlikely trials ($X^2_{1}=0.30$, $p=0.582$), but did affect performance on likely trials ($X^2_{1}=28.69$, $p<0.001$; Figure 3A). These results suggest that a likely sentence ending is less likely to be perceived as such when a prediction is harder to make. Trials with incorrect or missing responses were excluded from the neuroimaging analysis.

On the localizer tasks, participants performed well in all conditions (Figure 3B, average hits 92%, SEM = 2.3%, range 80-96%). The mixed-model ANOVA revealed significant effects of Condition ($F_{1,80}= 4.84$, $p = 0.03$), Task ($F_{2,80}=44.76$, $p<0.001$) and the interaction between Condition and Task ($F_{2,80} = 17.60$, $p<0.001$). Follow-up tests showed that these effects were driven by overall slightly poorer performance in the orthographic attention task than the other tasks (Orthographic vs. Phonological: $X^2_{1}=63.17$, $p<0.001$; Orthographic vs. Semantic: $X^2_{1}=70.88$, $p<0.001$), and poorer performance in the Orthographic 0-back task than the 1-back task ($X^2_{1}=36.01$, $p<0.001$). No significant differences were present in performance between the phonological and the semantic localizers, or between 1-back and 0-back conditions of these
tasks (Figure 3B). These results suggest that the orthographic (visuospatial) localizer was more difficult than the other two tasks.

Imaging results

Areas that respond to written meaningful language (reading contrast)

The reading contrast revealed a widespread network of cortical and subcortical regions that are classically implicated in language processing, attention and visual processing (Price, 2012; Rodd et al., 2015). Areas engaged when processing the sentence stem were bilateral inferior and middle frontal gyrus, medial frontal gyrus, bilateral middle temporal gyrus extending from the temporal pole into temporoparietal cortex, left thalamus, bilateral posterolateral cerebellum and the cerebellar vermis (Figure 3A, Table 1). Activations were more pronounced on the left of the cerebral cortex, and on the right in the cerebellum.

Areas where activity covaries with predictability (predictability contrast):

The predictability contrast revealed an area in right posterolateral cerebellum, Crus II, where haemodynamic activity positively correlated with predictability (Figure 4B, 5B, Table 1). Supratentorial clusters were identified in the left inferior frontal gyrus, right middle frontal gyrus, left posterior parietal
cortex, pre-supplementary motor area, and right caudate nucleus (Figure 4B, 6, Table 1). No brain areas showed activity that correlated negatively with the predictability of the items. All clusters apart from the right middle frontal gyrus cluster overlapped with a map of regions that are functionally connected to Crus I and Crus II (Figure 4C).

ROI analyses: cerebellar area that represents prediction also represent prediction error

A paired-samples t-test compared the regression weights for unlikely outcomes and likely outcomes for the cerebellar cluster that was modulated by predictability (predictability contrast) and for the cerebellar area that responded to written language (reading contrast). As only one cluster was identified in each contrast, tests were considered significant at p<0.05. The Crus II cluster that was modulated by predictability (Figure 5B) also showed a larger response to unlikely than to likely sentence outcomes (MNI 28 -86 -48 likely > unlikely: t_{16} = 2.27, p = 0.037). Conversely, the larger area that responded to the stem event (Figure 5A) did not show such a difference (MNI 30 -70 -52: t_{16}=0.33, p=0.743).
ROI analyses: cerebellar area that represents predictability is engaged in phonological processing but not semantic or orthographical processing.

Paired t-tests compared the activity in 0-back and 1-back conditions for semantic, visual and phonological localizers. This analysis indicates whether the areas that were modulated by predictability were also engaged by attention to semantic content, phonological or visual features. Results showed that right Crus II was significantly engaged in the phonological localizer task (MNI 28 -86 -48; $t_{15} = 2.52$, $p = 0.032$), but not in the semantic or orthographic task (Figure 5B). Note that these results do not imply that the Crus II region is more engaged in the phonological task as compared with the other two tasks. The condition effect (1-back minus 0-back) in the phonological task differs from that in the semantic task ($t_{15} = 2.49$, $p = 0.025$), but not from that in the orthographic task ($t_{15} = 1.23$, $p = 0.238$). However, these between-task comparisons do not survive a Bonferroni correction for the three possible post-hoc tests. The Crus II region that responded to written language (reading contrast), was significantly recruited in all three localizer tasks (MNI 30 -70 -52. Semantic: $t_{15} = 8.82$, $p<0.001$; Phonological: $t_{15} = 9.08$, $p<0.001$; Orthographic: $t_{15} = 7.43$, $p<0.001$). This activation pattern is consistent with a region that is engaged in processing written meaningful language, as this entails semantic, phonological and orthographic processing.
In summary, we found that a discrete region in cerebellar Crus II was significantly modulated by the predictability of the stem sentence, in the interval before the outcome was presented. This area was also active in a contrast that probed phonological processing, but not in contrasts that probed semantic or visual processing. It lay within a broader zone of the cerebellum activated by the reading task (but not modulated by predictability) and that broader zone did overlap with the regions activated by semantic and orthographic processing.

Discussion

The right posterior cerebellum is consistently implicated in language processing, but its precise contribution remains unclear. In parallel with the predictive function of cerebellar motor regions through internal models of movements (Courchesne and Allen, 1997; Miall, 1998; Ebner and Pasalar, 2008), internal model prediction may generalize to non-motor cerebellar regions, particularly Crus I/II (Ramnani, 2006; Ito, 2008). Thus, language-sensitive right cerebellar regions may assist linguistic processing by predicting upcoming sentence content. Here, we tested this hypothesis with a closely controlled event-
related fMRI study. We compared activity time-locked to the presentation of identical sentence fragments that varied in the degree to which they predicted the final word of the sentence (their cloze probability).

Crucially, this sentence fragment was modeled independently from a context sentence, through which predictability was manipulated, and from the final word (outcome) of the sentence. We were thus able to capture effects of prediction in the absence of outcome evaluation or prediction error, while also avoiding motor, linguistic, and working memory confounds. Further, using separate fMRI localizer tasks, we assessed whether identified prediction-sensitive areas were also engaged in semantic, phonological, or orthographic processing.

As hypothesized, activity in right Crus II increased with the predictability of the upcoming sentence ending. Further consistent with the presence of internal model predictions, the same Crus II area was more active during an unexpected outcome (prediction error) than an expected outcome. Finally, this area was also engaged when attending to phonological information, but not semantic or orthographic information.

This study is the first to identify a right cerebellar region that represents predictability independently from motor demands or error processing. Our findings complement and extend existing evidence on linguistic
prediction in the right posterolateral cerebellum. Previous fMRI evidence indicates that right posterior
cerebellar regions are engaged when a linguistic prediction is possible (Desmond et al., 1998; Moberget
et al., 2014). We have previously shown that low-frequency right cerebellar rTMS disrupts the prediction
of upcoming sentence content in a language comprehension task (Lesage et al., 2012), a finding we
recently replicated using cathodal tDCS (Miall et al., 2016). In the language production domain, rTMS
over right, but not left cerebellum impairs higher level speech monitoring – including internal prediction of
upcoming speech (Runnqvist et al., 2016), and a recent study found that right cerebellar tDCS improved
performance in a sentence completion task (D'Mello et al., 2017). Such neurostimulation evidence
dovetails nicely with the present data to show that the right posterior cerebellum is causally involved in
linguistic prediction to aid both language comprehension and language production.

A posterolateral cerebellar contribution to language processing is consistent with the region’s connectivity
Akkal et al., 2007; Bostan et al., 2013), resting-state functional connectivity and meta-analytic connectivity
mapping in humans (Habas et al., 2009; Krienen and Buckner, 2009; Buckner et al., 2011; Bernard et al.,
2012; Balsters et al., 2013) have identified connectivity between Crus I/II and higher-order cognitive and
language regions, including inferior frontal, dorsolateral prefrontal, posterior parietal, and anterior
cingulate cortices. In the present data, cerebral areas where the haemodynamic response scaled with linguistic predictability included the left inferior frontal gyrus, pre-SMA and left posterior parietal lobe, right middle frontal gyrus and bilateral caudate nucleus. These areas are all implicated in lexico-semantic or phonological language processing (Fedorenko et al., 2010; Wu et al., 2012; Martin et al., 2015), and all except the right DLPFC cluster were within the network of regions functionally connected to right Crus I/II (Bernard et al., 2012).

Our findings support the idea that cerebellar internal models aid language comprehension by predicting upcoming stimuli. Internal models are prominent in theories of motor cerebellar function (Miall, 1998; Wolpert et al., 1998), and it has long been hypothesized that cognitive and linguistic internal models could be present in prefrontal-projecting cerebellar areas (Leiner et al., 1989; Ramnani, 2006; Ito, 2008). Internal model prediction has been incorporated into psycholinguistic accounts more recently (Hickok, 2012; Rothermich and Kotz, 2013; Kotz et al., 2014; Pickering and Garrod, 2014). One fairly comprehensive theoretical frameworks posits that comprehension is achieved using the speech production apparatus, with both speech production and comprehension aided by internal model prediction (Pickering and Garrod, 2013; Pickering and Clark, 2014). This model aligns well with our present findings and previous neurostimulation and neuroimaging evidence (Lesage et al., 2012; Moberget et al., 2014;
Miall et al., 2016; D'Mello et al., 2017), which indicate that prediction of upcoming words may occur in or depend upon the cerebellum.

A major challenge in determining the function of prefrontal-projecting cerebellar areas is their involvement in processes that are difficult to manipulate separately. Notably, the Crus I/II area implicated in language is also consistently implicated in verbal working memory, where recruitment scales with cognitive load (Hayter et al., 2007; Lesage et al., 2010; Marvel and Desmond, 2010, 2013). Indeed, it has been proposed that the posterior cerebellum may act as the Baddeley and Hitch's (1974) phonological store, encoding verbal content and keeping this information online (Chen and Desmond, 2005; Marvel and Desmond, 2010). However, the involvement of the posterior cerebellum in language cannot be explained entirely by working memory demands. The right posterolateral cerebellum is recruited consistently in lexico-semantic processing (Vandenberghe et al., 1996; Fedorenko et al., 2010; Price, 2012; Lesage et al., 2015), even in relatively undemanding conditions, such as reading meaningful sentences as contrasted with more cognitively demanding scrambled sentences (Moberget et al., 2014). To explore functional overlap between working memory and language processes, we assessed cerebellar recruitment in three 1-back tasks that each captured a component of reading; attention to a semantics (semantic categorization), attention to phonology (rhyming judgment) or attention to orthographic features.
(visuospatial matching). In the present data, we found that the prediction-sensitive cerebellar cluster was engaged in the phonological task, but we did not find that this area was engaged in the semantic or orthographic tasks. This area’s recruitment in a phonological task aligns with a cerebellar role in the phonological store and inner speech (Ackermann et al., 2004, 2007, Marvel and Desmond, 2010, 2013). The absence of this area’s significant engagement in the semantic task is somewhat surprising, especially as evidence for cerebellar linguistic prediction is largely derived from semantic prediction tasks, including the task used here (Lesage et al., 2012; Argyropoulos, 2015; Miall et al., 2016; D’Mello et al., 2017). However, our data do not necessarily mean that internal models exclusively predict the phonological form of upcoming content, or that this prediction cannot be semantic. For example, semantic predictions may be represented in a common code to the representations needed in the phonological task. Alternatively, the semantic task, which used line drawings, may have captured semantic processes distinct from those in the prediction task, and a different localizer task might have recruited the prediction-sensitive cerebellar region.

A larger area of right Crus II that was consistently activated during reading (when meaningful language was presented) but not specific to prediction, was robustly engaged in all three localizer tasks. This is consistent with imaging evidence for semantic processing in posterolateral cerebellum (Price, 2012) and with meta-analyses of cerebellar recruitment in various tasks, where clusters responding to verbal
working memory and language tasks overlap (Stoodley and Schmahmann, 2009, 2010; Keren-Happuch et al., 2012; Stoodley et al., 2012).

This study is not without limitations. First, stimulus type differed between the localizer tasks. Even though the contrasts used controlled for such lower-level differences, it is possible that a semantic localizer using written language might have produced different results, potentially recruiting the cerebellar area that scaled with predictability. Second, the analysis on the localizer tasks is unable to speak to whether regions are recruited differently in different localizer tasks. Third, the order of the localizer task runs was not counterbalanced. We can therefore not exclude fatigue or learning effects. However, given the lack of performance differences, we think it unlikely that order affected the phonological or semantic localizer tasks. Lower performance in the 0-back condition of the orthographic localizer may be partially attributable to fatigue, but it is not clear what outcome such an order effect would have on cluster location.

Future research can further elucidate how working memory and linguistic prediction are represented in the cerebellum and whether internal model prediction could be an underlying mechanism to support these functions. Tasks using different stimulus types may further shed light on how linguistic prediction takes place in the cerebellum. Finally, study of the interaction between supratentorial areas that are functionally
connected to the cerebellum and also represent predictability, can elucidate how linguistic internal model prediction is achieved.

Conclusions

We identified an area in cerebellar Crus I/II where BOLD response scales with the predictability of upcoming sentence content. Activity in this region was larger when an unexpected sentence ending was evaluated compared to an expected sentence ending, consistent with processing prediction errors. Interestingly, the cerebellar area modulated by predictability was also recruited in a phonological processing task, but not in orthographic or semantic processing tasks. Thus, results support the presence of linguistic internal models during language comprehension and suggest that this process may rely on phonological processing.
References


Keren-Happuch E, Chen S-HA, Ho M-HR, Desmond JE (2012) A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. Hum Brain Mapp 0.


Figure 1. Trial structure of the prediction task. The stem and the outcome stimuli are matched for high- and low-cloze trials, and the context sentences are matched for length. Here, two items with the same stem, where one has a very predictable sentence ending (cloze probability 0.95) and the other does not (cloze probability 0.31). Three temporal events are independently modeled in the analysis: the context (3s), the stem (1s) and the outcome (1s). Ellipsis indicate variable delay/temporal jitter.
Figure 2. Trial structure and typical stimuli for the localizer tasks. A. Stimulus timing. B-D. The upper overlapping panels show typical stimulus displays (stimuli presented 500ms, 1s apart), illustrating first a match and then a non-match trial, for the 1-back runs. The separate lower panel shows the target item for the 0-back runs.
Figure 3. Behavioral performance. A. Percentage accurate responses in prediction task. Discrete levels of predictability were used for display purposes only; analyses were conducted using predictability as a continuous variable. B. Percentage accurate responses in localizer tasks. Error bars indicate +/- 1 SEM.

***: p<0.001.
Figure 4. Imaging results A. Areas activated over baseline when reading (reading contrast). B. Areas where BOLD response is modulated by predictability of future outcome (predictability contrast). Contrasts family-wise error corrected at alpha<0.05 (voxelwise p<0.001, cluster size > 99 voxels). C. Areas functionally connected to right Crus I and Crus II, based on Bernard et al. (2012) in red, with the results from the predictability contrast (yellow) overlaid to indicate overlap.
Figure 5: Parameter estimates for right cerebellar activations. First column: Cerebellar clusters in reading contrast (A, blue) and predictability contrast (B, yellow), whole-brain corrected at FWE p<0.05. Column 2 – Predictability. Parameter estimates extracted for different levels of predictability. Regression weights were extracted from the clusters identified in the imaging analysis where predictability (cloze probability) was a continuous variable; they are plotted to aid interpretation only – no statistical inference should be drawn. Column 3 – Outcome. ROI analysis for prediction error, using clusters as ROIs. Column 4 - Localizer task responses. ROI analysis for semantic, phonological and orthographic processing (1-back minus 0-back), using the same cluster masks. Paired t-tests. * p<0.05, *** p<0.001. a.u.: arbitrary units.

Error bars denote SEM.
Figure 6: Parameter estimates for cerebral areas engaged in prediction. Left column: Supratentorial brain areas that relate to predictability, whole-brain corrected at FWE p<0.05. Column 2 – Predictability. Parameter estimates extracted for varying levels of predictability. Regression weights were extracted from the clusters identified in the imaging analysis where predictability (cloze probability) was a continuous variable; they are plotted to aid interpretation only – no statistical inference should be drawn. Column 3 – Outcome. Parameter estimates for prediction errors; as these areas were not part of an a priori hypothesis, no statistical inference should be drawn. Column 4 – Localizers. Parameter estimates (1-back minus 0-back) for semantic, phonological and orthographic processing; again, as these areas were not part of an a priori hypothesis, no statistical inference should be drawn. a.u.: arbitrary units. Error bars denote SEM.
Table 1. Table of results. Cluster corrected (FWE corrected alpha <0.05: voxel-wise p<0.001, clusters size > 99 voxels). For clusters that encompass multiple peaks, the volume of the entire cluster is given, with the index of the sub-peak in parenthesis.

<table>
<thead>
<tr>
<th>Gross anatomical region</th>
<th>volume (mm³)</th>
<th>T</th>
<th>MNI coordinates</th>
<th>Cytoarchitectonic region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STEM</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Frontal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Inferior Frontal Gyrus</td>
<td>179,384 (5)</td>
<td>10.57</td>
<td>-50</td>
<td>12</td>
</tr>
<tr>
<td>Left Precentral Gyrus</td>
<td>179,384 (6)</td>
<td>9.87</td>
<td>-50</td>
<td>-6</td>
</tr>
<tr>
<td>Right Inferior Frontal Gyrus</td>
<td>32,440 (1)</td>
<td>6.03</td>
<td>48</td>
<td>18</td>
</tr>
<tr>
<td>Right Middle Frontal Gyrus</td>
<td>32,440 (2)</td>
<td>8.48</td>
<td>50</td>
<td>2</td>
</tr>
<tr>
<td>Left Superior Frontal Gyrus</td>
<td>17,592</td>
<td>13.19</td>
<td>-6</td>
<td>8</td>
</tr>
<tr>
<td><strong>Parietal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Inferior Parietal Lobule</td>
<td>12,608</td>
<td>8.94</td>
<td>32</td>
<td>-52</td>
</tr>
<tr>
<td>Left Inferior Parietal Lobule</td>
<td>12,472</td>
<td>6.21</td>
<td>-34</td>
<td>-58</td>
</tr>
<tr>
<td><strong>Occipital</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Middle Occipital Gyrus</td>
<td>179,384 (1)</td>
<td>14.69</td>
<td>-36</td>
<td>-94</td>
</tr>
<tr>
<td>Left Inferior Occipital Gyrus</td>
<td>179,384 (3)</td>
<td>11.98</td>
<td>-44</td>
<td>-60</td>
</tr>
<tr>
<td>Right Inferior Occipital Gyrus</td>
<td>118,064 (1)</td>
<td>12.25</td>
<td>42</td>
<td>-92</td>
</tr>
<tr>
<td>Right Inferior Occipital Gyrus</td>
<td>118,064 (3)</td>
<td>11.3</td>
<td>28</td>
<td>-94</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Inferior Temporal Gyrus</td>
<td>179,384 (2)</td>
<td>13.14</td>
<td>-40</td>
<td>-46</td>
</tr>
<tr>
<td>Left Middle Temporal Gyrus</td>
<td>179,384 (4)</td>
<td>11.65</td>
<td>-54</td>
<td>-50</td>
</tr>
<tr>
<td>Right Inferior Temporal Gyrus</td>
<td>118,064 (2)</td>
<td>12.56</td>
<td>44</td>
<td>-62</td>
</tr>
<tr>
<td>Right Middle Temporal Gyrus</td>
<td>118,064 (4)</td>
<td>9.99</td>
<td>54</td>
<td>-36</td>
</tr>
<tr>
<td>Right Superior Temporal Gyrus</td>
<td>118,064 (6)</td>
<td>6.92</td>
<td>60</td>
<td>2</td>
</tr>
<tr>
<td>Left Temporal Pole</td>
<td>179,384 (7)</td>
<td>9.33</td>
<td>-52</td>
<td>10</td>
</tr>
<tr>
<td>Left Middle Temporal Gyrus</td>
<td>179,384 (8)</td>
<td>8.99</td>
<td>-56</td>
<td>-6</td>
</tr>
</tbody>
</table>

**Insular regions**
Right Insula Lobe 2,552 6.15 34 24 4

Cerebellum
Right cerebellum 118,064 (5) 9.59 30 -70 -52 Lobule VIIb (Hem)
Right cerebellum 118,064 (7) 6.37 30 -62 -26 Lobule VI (Hem)
Left cerebellum 1,304 5.68 -30 -70 -52 Lobule VIIb (Hem)

Other subcortical
Left Thalamus 840 4.55 -8 -16 12

STEM COVARIATE

Frontal
Left Superior Frontal Gyrus 1,584 4.43 0 28 62 BA8/pre-SMA
Left Inferior Frontal Gyrus 1,592 4.14 -42 22 -10 BA47
Right Middle Frontal Gyrus 1,360 4.66 44 20 40 BA 9/46

Parietal
Left Superior parietal lobule 1,600 4.04 -50 -58 56 BA 7

Cerebellum
Right Cerebellum 1,072 4.19 28 -86 -48 Lobule VIIa Crus II (Hem)

Other subcortical
Right caudate nucleus 2,664 4.76 6 4 18