

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

Investigating cortico-cerebellar contributions to cognition

### Permalink

<https://escholarship.org/uc/item/61w6t6sk>

### Author

King, Maedbh

### Publication Date

2022

Peer reviewed|Thesis/dissertation

Investigating cortico-cerebellar contributions to cognition

By

Maedbh King

A dissertation submitted in partial satisfaction of the  
requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Richard Ivry, Chair  
Professor Silvia Bunge  
Professor Kevin Weiner

Spring 2022

Investigating cortico-cerebellar contributions to cognition

Copyright 2022  
by  
Maedbh King

## Abstract

Investigating cortico-cerebellar contributions to cognition

by

Maedbh King

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Richard Ivry, Chair

Anatomical, neuropsychological, and neuroimaging research has revealed extensive connectivity between most of the cerebral cortex and the cerebellum, pointing to the involvement of the human cerebellum in a broad range of functions beyond its traditional association with motor control. In a series of studies, I used functional magnetic resonance imaging, neuropsychology, and behavior to investigate how cortico-cerebellar connectivity contributes to cognition. In Chapter 1, I leveraged a multi-domain task battery to develop a comprehensive functional map of the human cerebellum, and in Chapter 2, I adopted a machine learning approach to build models of cortico-cerebellar connectivity to determine how this functional map is constrained by cortical input patterns. In Chapter 3, I showed that there is no selective impairment of the cerebellum in semantic prediction processing, and in Chapter 4, I demonstrated that the architecture of cortico-cerebellar connectivity changes over the course of learning for dynamic action prediction tasks. I hope that findings from this work can be incorporated by future studies to better understand how cortico-cerebellar networks are impacted in neurological disease and developmental disorders that affect the cerebellum.

## Dedication

This work is dedicated to the underappreciated, and oft maligned, “little brain”, and all the scientists whose work inspired me to contribute, in a small way, to the “cerebellar cognitive revolution”.

# Acknowledgements

In my native Irish tongue there is a saying, *“Ar scáth a chéile a mhaireann na daoine”*, which directly translates to “people live in each other’s shadows”. Throughout this journey, I have lived “in the shadows” of a great many people, and this work would not have been possible without their support.

Pursuing a PhD is an exercise in humility, resilience, and a kind of absurdity. I am grateful to my fellow lab members, past and present, for creating a collegial, fun and productive environment in which to pursue science.

Sometimes you are lucky enough that one of your lab members becomes a treasured friend. Christina, I am so fortunate that you accepted me into your Berkeley community when I first moved here, and have continued to be a loyal and fast friend.

Susan, Ashley, and Juliana, I’m thankful to each one of you for your enduring friendship and for sharing California sunsets, rooftop drinks, and all other occasions, big and small, with me.

Manon, thank you for being my soul-friend. I trust you implicitly, because I know that you will always tell me the truth.

To the friends who have known me since childhood, especially Róisín and Kathryn, thank you for the support you have given me from afar over the years. You always made the effort to stay connected with me, and invited me back into the fold whenever I returned to Ireland.

To my parents, Catherine and Tim, thank you for supporting me in all of my endeavors. You always believed in me and allowed me to follow my instincts, whether or not you understood my motivations (I don’t know that I always did). The confidence I have in my own ability is a product of the secure and happy environment you cultivated for me growing up. To my siblings, Dara and Deirbhile, I have lived abroad for much of our adult lives, and I’m sorry that we haven’t been able to see each other as much as I would have liked. I hope to change that in the future.

To my mentor, Rich. A PhD advisor can make or break the journey, and I am thankful everyday that you made my time at Berkeley so worthwhile and meaningful. It has been a privilege to learn under your tutelage over the years. You have shown me that it is

possible to be everything at once: a successful scientist, a kind mentor, and a well-rounded, interesting person.

Finally, to my partner, Matthew. We met the summer after the first year of my PhD and you have been by my side ever since. I want to thank you for providing me with enormous practical and technical support, from encouraging me to switch from MATLAB to Python early on, and for being patient with me as I struggled to understand the nuances of object-oriented programming. A renaissance man, you have always been a major source of intellectual inspiration: your encyclopedic knowledge along with your measured approach to thinking about complex ideas has challenged me to reconsider my own preconceptions, and to think deeply about *what* and *how* I want to contribute to the world, and *who* I want to be. I have asked a lot of you throughout this journey, challenging you in more ways than one, but you have always and continue to be, patient and present and above all else, kind. You see me exactly as I am, and you have expanded my mind in ways that a PhD could not. I know how fortunate I am to have met you, and I could not imagine a better partner-in-life than you.

# Table Of Contents

|                                                                       |           |
|-----------------------------------------------------------------------|-----------|
| <b>Abstract</b>                                                       | <b>1</b>  |
| <b>Dedication</b>                                                     | <b>i</b>  |
| <b>Acknowledgements</b>                                               | <b>ii</b> |
| <b>Table Of Contents</b>                                              | <b>iv</b> |
| <b>General Introduction</b>                                           | <b>1</b>  |
| <b>Chapter 1: Functional Organization of the Human Cerebellum</b>     | <b>7</b>  |
| 1.1 Abstract                                                          | 7         |
| 1.2 Introduction                                                      | 7         |
| 1.3 Methods                                                           | 8         |
| 1.3.1 Participants                                                    | 8         |
| 1.3.2 Experimental tasks                                              | 8         |
| 1.3.3 Experimental design                                             | 9         |
| 1.3.4 Hand assignment across tasks                                    | 10        |
| 1.3.5 Behavioral training                                             | 10        |
| 1.3.6 Eye tracking                                                    | 11        |
| 1.3.7 Scanning sessions                                               | 11        |
| 1.3.8 Image acquisition                                               | 11        |
| 1.3.9 Image preprocessing                                             | 12        |
| 1.3.10 General linear model (GLM)                                     | 12        |
| 1.3.11 Cerebellar spatial normalization                               | 13        |
| 1.3.12 Motor feature model                                            | 13        |
| 1.3.13 Reliability of activation patterns                             | 14        |
| 1.3.14 Spatial frequency of activation patterns                       | 14        |
| 1.3.15 Evaluating functional boundaries                               | 14        |
| 1.3.16 MDTB data set parcellation                                     | 15        |
| 1.3.17 Bootstrap analysis                                             | 16        |
| 1.3.18 Evaluation of functional parcellations                         | 16        |
| 1.3.19 Representational structure of task-related activation patterns | 17        |
| 1.3.20 Feature-based approach                                         | 18        |
| 1.3.21 Statistical analysis and inference                             | 18        |
| 1.4 Results                                                           | 18        |
| 1.4.1 Identification of motor features from the MDTB.                 | 19        |

|                                                                                                                                                 |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 1.4.2 MDTB elicits varied activation patterns across the cerebellum.                                                                            | 20        |
| 1.4.3 Cerebellar lobules do not reflect functional subdivisions.                                                                                | 21        |
| 1.4.4 MDTB parcellation uncovers strong functional boundaries.                                                                                  | 22        |
| 1.4.5 Task-free parcellations identify overlapping but weaker boundaries.                                                                       | 25        |
| 1.4.6 Characterizing activation by cognitive features.                                                                                          | 27        |
| 1.5 Discussion                                                                                                                                  | 28        |
| 1.5.1 Summary                                                                                                                                   | 28        |
| 1.5.2 Parcellations of the cerebellar cortex.                                                                                                   | 29        |
| 1.5.3 Insights about functional topography.                                                                                                     | 30        |
| 1.5.4 Conclusions                                                                                                                               | 31        |
| 1.6 Supplementary Material                                                                                                                      | 32        |
| <b>Chapter 2: A task-general connectivity model reveals variation in convergence of cortical inputs to functional regions of the cerebellum</b> | <b>40</b> |
| 2.1 Abstract                                                                                                                                    | 40        |
| 2.2 Introduction                                                                                                                                | 40        |
| 2.3 Results                                                                                                                                     | 42        |
| 2.3.1 Overview                                                                                                                                  | 42        |
| 2.3.2 Cortico-cerebellar connectivity is best captured by models with convergence                                                               | 43        |
| 2.3.3 Lobules VII receives more distributed input from the cerebral cortex                                                                      | 45        |
| 2.3.4 Cortico-cerebellar connectivity model predicts new cerebellar data                                                                        | 48        |
| 2.4 Discussion                                                                                                                                  | 49        |
| 2.4.1 Convergence differs across Cerebellar Circuits                                                                                            | 50        |
| 2.4.2 Task-based vs. resting-state connectivity analyses                                                                                        | 50        |
| 2.4.3 Methodological limitations                                                                                                                | 51        |
| 2.4.4 Conclusions                                                                                                                               | 52        |
| 2.5 Methods                                                                                                                                     | 52        |
| 2.5.1 Multi-domain task battery                                                                                                                 | 52        |
| 2.5.2 Image acquisition and preprocessing                                                                                                       | 53        |
| 2.5.3 General Linear Model (GLM)                                                                                                                | 53        |
| 2.5.4 Cerebral Cortex Surface Reconstruction                                                                                                    | 53        |
| 2.5.5 Spatial Normalization of Cerebellar Data                                                                                                  | 53        |
| 2.5.6 Connectivity Models                                                                                                                       | 54        |
| 2.5.7 Cortical parcels                                                                                                                          | 54        |
| 2.5.8 Model Estimation                                                                                                                          | 55        |
| 2.5.9 Model Testing                                                                                                                             | 55        |
| 2.5.10 Noise Ceiling                                                                                                                            | 56        |
| 2.5.11 Cortico-Cerebellar Convergence                                                                                                           | 57        |

|                                                                                                           |           |
|-----------------------------------------------------------------------------------------------------------|-----------|
| 2.5.12 Generalization to new participants                                                                 | 57        |
| 2.6 Supplementary Figures                                                                                 | 58        |
| <b>Chapter 3: No evidence for semantic prediction deficits in patients with cerebellar degeneration</b>   | <b>65</b> |
| 3.1 Abstract                                                                                              | 65        |
| 3.2 Introduction                                                                                          | 65        |
| 3.3 Methods                                                                                               | 69        |
| 3.3.1 Construction and Evaluation of Stimuli                                                              | 69        |
| 3.3.2 Participants (Main Experiment)                                                                      | 71        |
| 3.3.3 Procedure                                                                                           | 74        |
| 3.4 Results                                                                                               | 75        |
| 3.5 Discussion                                                                                            | 78        |
| 3.5.1 Conclusions                                                                                         | 81        |
| 3.6 Supplementary Figures                                                                                 | 82        |
| <b>Chapter 4: Unique cerebellar contribution to early learning: evidence from action prediction tasks</b> | <b>84</b> |
| 4.1 Abstract                                                                                              | 84        |
| 4.2 Introduction                                                                                          | 84        |
| 4.3 Methods                                                                                               | 88        |
| Participants                                                                                              | 88        |
| 4.3.1 Experimental Tasks                                                                                  | 88        |
| 4.3.2 Construction and Evaluation of Stimuli: Action Prediction Task                                      | 89        |
| 4.3.3 Construction and Evaluation of Stimuli: Social Prediction Task                                      | 90        |
| 4.3.4 Experimental Design                                                                                 | 91        |
| 4.3.5 Scanning Sessions                                                                                   | 92        |
| 4.3.6 Behavioral Training                                                                                 | 92        |
| 4.3.7 Image Acquisition                                                                                   | 93        |
| 4.3.8 Anatomical data preprocessing                                                                       | 93        |
| 4.3.9 Functional data preprocessing                                                                       | 94        |
| 4.3.10 General Linear Model (GLM)                                                                         | 95        |
| 4.3.11 Cerebral Cortex Surface Reconstruction                                                             | 95        |
| 4.3.12 Spatial Normalization of Cerebellar Data                                                           | 95        |
| 4.3.13 Connectivity Models                                                                                | 96        |
| 4.3.14 Cortical parcels                                                                                   | 96        |
| 4.3.15 Model Testing                                                                                      | 97        |
| 4.3.16 Reliability of cerebellar activity patterns                                                        | 97        |
| 4.3.17 Evaluating task-specific deviations from the model                                                 | 97        |
| 4.4 Results                                                                                               | 97        |

|                                                                                                                             |            |
|-----------------------------------------------------------------------------------------------------------------------------|------------|
| 4.4.1 Overview                                                                                                              | 97         |
| 4.4.2 Participants learn cognitive tasks across multiple weeks                                                              | 99         |
| 4.4.3 Cortico-Cerebellar Connectivity Model Predicts New Tasks                                                              | 100        |
| 4.4.4 Cerebellar BOLD signal decreases across learning                                                                      | 102        |
| 4.4.5 Early learning is underpredicted by a model of cortico-cerebellar connectivity, but only for dynamic prediction tasks | 103        |
| 4.5 Discussion                                                                                                              | 106        |
| 4.5.1 Summary                                                                                                               | 106        |
| 4.5.2 Behavioral changes across learning                                                                                    | 106        |
| 4.5.3 Unique cerebellar contribution to dynamic action prediction tasks                                                     | 107        |
| 4.5.4 Limited evidence that the cerebellum plays a unique role in semantic prediction                                       | 108        |
| 4.5.5 Cerebellum plays a task-specific role in coordinating cortico-cortico interactions during early learning              | 108        |
| 4.5.6 Conclusions                                                                                                           | 109        |
| 4.6 Supplementary Figures                                                                                                   | 110        |
| <b>General Discussion</b>                                                                                                   | <b>116</b> |
| <b>6.1 References</b>                                                                                                       | <b>118</b> |
| 6.1.1 Chapter 1                                                                                                             | 118        |
| 6.1.2 Chapter 2                                                                                                             | 121        |
| 6.1.3 Chapter 3                                                                                                             | 124        |
| 6.1.4 Chapter 4                                                                                                             | 128        |
| 6.1.5 General Introduction and Discussion                                                                                   | 131        |

# General Introduction

The function of the cerebellum has been the focus of significant debate and empirical investigation over the past three decades. The cerebellum is positioned in the rostral roof of the fourth ventricle (Voogd & Glickstein, 1998), and although containing only 10% of the volume of the brain, houses approximately 60% of the total number of neurons in the human brain (Van Bartheld et al. 2016; Sereno et al. 2020). Anatomically, the cerebellum receives input from nearly the entire cerebral cortex (Schmahmann & Pandya, 1997) making it the largest bandwidth of information in the human brain (Kelly & Strick, 2003).

Historically, this structure has been associated with sensorimotor control and learning (Holmes, 1917, 1939), however, since the emergence of cognitive neuroscience in the 1980's, a more comprehensive picture has come to prominence, one implicating the cerebellum in functions that extend well beyond the motor domain (Stoodley & Schmahmann, 2009; Strick et al. 2009; Moberget & Ivry, 2016). Evidence for a cerebellar involvement in cognitive processes stems from a three-pronged approach: neuroanatomy, neuropsychology, and neuroimaging. Anatomically, the seminal paper of Leiner, Leiner, and Dow (1986) observed a parallel expansion of lateral hemispheres of the primate cerebellar cortex and association regions of the cerebral cortex, prompting a "thought experiment" that the cerebellum may underlie higher-order cognitive processes. Simultaneously, a rich neuropsychology literature observed selective deficits on perceptual and cognitive tasks in patients with degeneration and focal lesions of the cerebellum (Ivry & Keele, 1989; Ivry & Diener, 1991; Helmuth et al. 1997; Schmahmann & Sherman, 1998; Casini & Ivry, 1999). The advent of neuroimaging marked an important catalyst in the expansion of the "cerebellar cognitive revolution" (Moberget & Ivry, 2016). Thousands of functional magnetic resonance imaging (fMRI) studies reported loci of cerebellar activation during a large number of cognitive tasks, even when experimental conditions were controlled for overt motor responses. Petersen et al. (1989) pioneered this neuroimaging revolution when they demonstrated that the right posterolateral cerebellum was activated in response to a semantic retrieval task, a decidedly non-motor task. Since then, cerebellar activation has been reported in studies of attention (Desmond & Fiez, 1998), language (Moberget et al. 2014; Lesage et al. 2016), social cognition (Van Overwalle et al. 2014), and many other cognitive domains. Thus, it is well established by now that the cerebellum is activated by a range of cognitive tasks, along with its better known role in sensorimotor control and learning (Wolpert et al. 1998; Shadmehr et al. 2010).

Neuroimaging, specifically fMRI, has been unparalleled in its ability to observe whole brain activation. However, its utility in understanding the unique contribution of the cerebellum to cognition has been limited. One major shortcoming of current fMRI research is the piecemeal approach that studies have traditionally used to study the cerebellum's involvement in cognition. Namely, studies have largely focused on a single task or domain as a way of investigating functional organization, and they have done this by mapping task-evoked activation of a single task to the cerebellar cortex. This

approach makes sense when you consider that most researchers are interested in understanding the neural substrates of specific cognitive processes, and this piecemeal approach allows them to gain this understanding by generating closely matched experimental and control task conditions. However, this restricted approach presents a challenge if the aim of cerebellar research is to characterize a comprehensive and broad functional organization across subregions of the cerebellum. Moreover, the current approach of localizing function to cerebellar sub-regions relies on macroanatomy, using the traditional lobular nomenclature, with the 10 lobules (I-X) organized along the superior to inferior axis (Larsell, 1947). However, at a structural level, lobulation is a crude measure of cerebellar functional compartmentation (White & Sillitoe, 2013). From a structural perspective, lobules allow for increased surface area and cellular packing density, and this is especially notable in the cerebellum given the number of granule cells that must be packaged into the developing cerebellum (White & Sillitoe, 2013). fMRI activations are traditionally localized to specific lobules, however, it is not at all clear whether the boundaries that demarcate functional subregions align with these lobular boundaries. Indeed, resting-state functional connectivity has suggested that there are prominent functional differences along the medial-to-lateral axis of a single lobule, and that functional networks span many lobules (Buckner et al. 2011; Diedrichsen & Zotow, 2015). Thus, there remain two unanswered questions: how should we localize task-evoked activation to the cerebellar cortex, and what is the compartmentation of functional subregions of the human cerebellum.

In Chapter 1, I set out to answer these questions. I did this by adopting a novel approach, one that offers a dramatic departure from existing paradigms. Specifically, I used a broad base, multi-domain task battery (King et al. 2019), spanning language, memory, attention, visual perception, social cognitive domains, to develop a novel functional parcellation of the human cerebellum. Moreover, I developed a novel evaluation criterion to determine the true functional boundaries of the human cerebellum. The functional parcellation was created using an unsupervised clustering approach, and to complement this bottom-up approach, I used a feature-based model to characterize activity across the cerebellum in terms of latent cognitive processes. I used the evaluation metric to compare the task-based parcellation with existing parcellations in the current literature, one based on resting-state fMRI connectivity, and the other based on macroanatomical structure. I also estimated the map on individual parcellations and quantified the degree to which they diverged from the normative group parcellation. Given our understanding of the anatomical and functional connections between the cerebral cortex and the cerebellum, in Chapter 2, I set out to explore the pattern of functional mapping from the cerebral cortex to distinct functional subregions of the cerebellum.

We know that there is a tight interplay between the cerebellum and the cerebral cortex: the cerebral cortex projects to the cerebellum via the pontine nuclei, and receives inputs from the cerebellum via the dentate nucleus and the thalamus (Dum & Strick, 2003; Middleton & Strick, 2001). The cerebro-ponto-cerebellar-thalamo system is organized in a highly stereotyped manner, and communication between the cerebral cortex and the cerebellum is often understood as relying on a series of parallel loops,

each connecting a single cortical and cerebellar region in a closed-loop circuit via pontine nuclei, deep cerebellar nuclei and the thalamus (Schmahmann & Pandya, 1997). For example, transsynaptic tracing work has shown that if you inject a retrograde tracer into Brodmann area 46 of the macaque (analogous to prefrontal cortex in humans), cells in the dentate nucleus are labeled, and transsynaptic tracers injected into the primary motor cortex label lobules IV-VI and lobule VIII, known somatotopic regions of the cerebellum (Kelly and Strick, 2003). Importantly, dentate neurons that are labeled after virus injections into prefrontal areas are located in regions spatially separate from those labeled after injections into motor areas of the cerebral cortex, suggesting that the distinct motor/non-motor functional topography is preserved across all levels of the circuit (Middleton & Strick, 2001; Dum & Strick, 2003). More recently, this tight interplay between the cerebral cortex and the cerebellum has been demonstrated using resting-state fMRI connectivity (O'Reilly et al. 2010; Buckner et al. 2011; Guell et al. 2018). For example, Buckner et al. (2011) generated a map of the human cerebellum by regressing each cerebellar voxel against cortical resting-state networks, and then did a winner-take-all approach to assign each voxel to the cortical network with which it was most correlated. The map they produced showed supporting evidence for connectivity between motor regions of the cortex and the cerebellum.

Given this architecture between the cortex and the cerebellum, insight into the functional activation of the cerebellar cortex necessitates that we simultaneously consider outputs from the cerebral cortex. Surprisingly, most fMRI studies investigating the cerebellum's role in cognition do not consider the relative cortical contributions to cerebellar function. This is especially problematic when we consider that the BOLD signal in the cerebellum is largely reflective of mossy fiber inputs (Mathiesen et al. 2000), which largely carry input from the cerebral cortex via the pontine nuclei (Llinás, 1975). Therefore, it is critical to examine, in parallel, the relationship between activation in the cerebral cortex and cerebellum, in order to meaningfully intercept cerebellar function.

Thus, the goal of Chapter 2 was to establish a quantitative model of connectivity between the cerebral cortex and the cerebellum in order to better understand domain-general functions of the cerebellum. In contrast to the cerebral cortex, where both input and spiking activity contribute to the BOLD signal (Logothetis et al. 2001; Attwell et al. 2002), the cerebellar BOLD signal provides a relatively pure measure of the input (Mathiesen et al. 2000). Consequently, the model focuses on explaining cerebellar activity as a function of cerebral activity (not the other way around). Using a large multi-domain task battery (King et al. 2019), I estimated connectivity models that could predict the cerebellar BOLD response as a function of activity patterns in the cerebral cortex. Previous neuroimaging work has suggested a one-to-one mapping between the cerebral cortex and the cerebellum (Buckner et al. 2011), however, I was interested in understanding the degree to which there is convergence from cortical regions onto the cerebellar cortex. The idea of convergence has been proposed in recent rodent work demonstrating that multiple cortical inputs converge in the cerebellar cortex, sometimes onto the same granule cell (Huang et al. 2013). This architecture suggests an information integration role between the cerebellum and distinct regions of

the cerebral cortex (Henschke et al. 2020). To uncover robust and reliable patterns of cortico-cerebellar connectivity, I trained models on cortical activity patterns and tested them on held-out cerebellar activity patterns. I varied the degree of convergence between the cerebral cortex and cerebellum in order to adjudicate between different possible architectures. Once a robust model had been validated, I examined differences in convergence from regions of the cerebral cortex onto distinct cerebellar sub-regions. In Chapter 3, I used this model to test for characteristic violations of cortico-cerebellar connectivity as a way of evaluating functional and mechanistic theories of the cerebellum.

A critical question of cognitive neuroscience research centers on what the cerebellum contributes to neural processing. Over the past three decades, a large number of theories have been proposed in an attempt to explain cerebellar contributions to cognition. Because this structure has historically been associated with sensorimotor control and learning (Holmes, 1917, 1939), many of the theories have been developed to explain motor control and sensorimotor function (Marr & Thach, 1969; Albus, 1970; Wolpert et al., 1995; Ito, 2008). Functional theories of the cerebellum have also been influenced by the finding that cytoarchitecture is largely homogenous across the cerebellar cortex (Eccles et al., 1967; Apps & Hawkes, 2009), suggesting that there is a basic computational function of the cerebellum (“Universal Cerebellar Transform” or UCT), which is repeated across domains as diverse as language, attention, motor control and social cognition (Schmahmann et al. 1996).

There are a number of prominent functional theories of the cerebellum, most notable are computational theories that center on the cerebellum’s pivotal role in generating error-based predictions (Criscimagna-hemminger et al. 2010; Shadmehr et al. 2010; Schlerf et al. 2012; Peterburs et al. 2019;), representing temporal relationships (Ivry & Keele, 1989), and performing internal simulations (Wolpert et al. 1998). More broadly, there are mechanistic theories that concern how the cerebellum may work in concert with the cerebral cortex to support specific cortical processes, rather than contributing unique computational functions such as model-based predictions and timing (Sokolov & Ivry, 2017). For example, the cerebellum may assist prefrontal cortical regions in processing rule-related representations (Miller and Cohen, 2001; Wallis et al., 2001), or the temporal parietal junction in interpreting social interactions (Saxe & Wexler, 2005). One mechanistic account of cortico-cerebellar communication is that the cerebellum works to ensure “mental coordination” of information within cortical networks (Andreasen & Pierson, 2008). Evidence for this idea often comes from dysfunction in cortico-cerebellar circuits underlying disorders like schizophrenia (Andreasen et al. 1998) and autism (Rogers et al. 2013). Symptoms of this dysfunction are distinctly cognitive in nature, including perseveration and inflexible thinking (Antonova et al. 2004).

Functional and mechanistic theories of the cerebellum have been studied using a variety of approaches, from neuroimaging to neurostimulation, and neuropsychology. Despite significant empirical efforts, it is still unclear what unique function(s) are contributed by the cerebellum to cognition. Therefore in Chapters 3 and 4, I used

neuropsychology and neuroimaging to address functional and mechanistic theories of the cerebellum's role in cognitive processing.

In Chapter 3, I used neuropsychology to test two functional theories of cerebellar contribution to semantic processing. Neuropsychology is a powerful approach that has been used to study how and what the cerebellum contributes to cognition by evaluating cognitive deficits in patients with cerebellar degeneration (Kansal et al. 2017), and cerebellar lesions (Ivry et al. 1988). There have been many studies that have identified deficits in a diverse range of cognitive tasks including prediction of temporal events (Breska & Ivry, 2018), mathematical reasoning (McDougle et al. 2022), and attentional processing (Baier et al. 2010). There is one particular domain that has received a lot of attention from both neuroimaging and neuropsychology approaches, and that is aspects of linguistic processing such as verb generation (Petersen et al. 1989) and semantic prediction (Fiez, 1992; Moberget et al. 2014). However, this domain presents a puzzle: there are many neuroimaging studies that identify a cerebellar involvement in some aspects of linguistic processing (Moberget et al. 2014; Lesage et al. 2017; D'Mello et al. 2017), however, there are also plenty of neuropsychology studies that do not find any evidence for cerebellar involvement in these same processes (Helmuth et al. 1997; Riva, 1998). To disentangle a possible cerebellar contribution to language, I took a neuropsychological approach and put a prominent theory of cerebellar function, the internal model hypothesis, to the test. Using a sentence verification task, I asked if individuals with cerebellar degeneration showed reduced sensitivity to semantic prediction compared with healthy controls. I 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.

In Chapter 4, I proposed a multi-session neuroimaging experiment that evaluates whether the cerebellum contributes a unique function to specific cognitive tasks, above and beyond what is predicted by activity patterns in the cerebral cortex. I did this by examining task-specific violations of a pre-trained cortico-cerebellar connectivity model across three different stages of learning. As a theory, learning has been made most salient by the seminal theoretical work of Marr and Thach (1969), and Albus (1971), who hypothesized that the simultaneous stimulation of mossy fiber and climbing fiber inputs arriving to the cerebellar cortex provides the perfect neural basis for error-based learning. The general principles underlying cerebellar learning can also be operationalized as a mechanistic function of the cerebellum. Namely, it can be understood as a way for the cerebellum to quickly establish new patterns of communication between regions of the cerebral cortex early in the learning process. Once established, the cerebellum may “drop out”, having taught the cerebral cortex the correct configuration for a given process. Using a model of cortico-cerebellar connectivity, I evaluated whether activity patterns in the cerebellum were underpredicted by the cerebral cortex early in the learning process, and whether this pattern changed over the course of learning. Both of these approaches, neuropsychological and neuroimaging, allowed me to determine whether the cerebellum makes unique functional and/or mechanistic contributions to cognition.

To summarize, in Chapter 1, I leveraged a multi-domain task battery to develop a comprehensive parcellation of the human cerebellum. In Chapter 2, I used the same battery of tasks to build and validate models of cortico-cerebellar connectivity. In Chapter 3, I used neuropsychology to test whether individuals with cerebellar degeneration show reduced sensitivity to semantic prediction, and in Chapter 4, I used neuroimaging to determine whether the cerebellum makes a unique contribution to cognition across learning.

# Chapter 1: Functional Organization of the Human Cerebellum

## 1.1 Abstract

There is compelling evidence that the human cerebellum is engaged in a wide array of motor and cognitive tasks. A fundamental question centers on whether the cerebellum is organized into distinct functional subregions. To address this question, we employed a rich task battery designed to tap into a broad range of cognitive processes. During four functional MRI sessions, participants performed a battery of 26 diverse tasks comprising 47 unique conditions. Using the data from this multi-domain task battery, we derived a comprehensive functional parcellation of the cerebellar cortex and evaluated it by predicting functional boundaries in a novel set of tasks. The new parcellation successfully identified distinct functional subregions, providing significant improvements over existing parcellations derived from task-free data. Lobular boundaries, commonly used to summarize functional data, did not coincide with functional subdivisions. The new parcellation provides a functional atlas to guide future neuroimaging studies.

## 1.2 Introduction

Converging lines of research provide compelling evidence that the cerebellum is engaged in a broad range of cognitive functions, well beyond its historical association with sensorimotor control<sup>1</sup>. Anatomical tracing studies in non-human primates have revealed reciprocal connections with parietal and prefrontal association cortices<sup>2</sup>. Individuals with lesions to the cerebellum exhibit behavioral impairments on tasks designed to assess non-motor processes such as duration discrimination, attentional control, spatial cognition, emotion perception and executive and language function. Perhaps most intriguing, neuroimaging studies consistently reveal activations of the cerebellar cortex during a diverse set of motor, cognitive, and social and affective tasks<sup>3</sup>. This raises the question of whether the cerebellum can be meaningfully subdivided into a discrete set of regions, reflecting distinct functional contributions across diverse task domains. In contrast to the cerebral cortex, the cytoarchitectonic organization is remarkably uniform across the entire cerebellar cortex. Due to this homogeneity, neuroimaging and neuropsychological studies have mostly relied on the macroanatomical folding of the cerebellum along the superior to inferior axis into ten lobules (numbered I–X)<sup>4</sup>. More recently, functional parcellations based on task-free functional magnetic resonance imaging (fMRI) data have been proposed<sup>5–7</sup>. However, the degree to which these proposed boundaries correspond to functional divisions remains unclear. Task-based studies have been limited by the lack of a comprehensive neuroimaging data set. A few studies have employed data sets involving multiple tasks<sup>7,8</sup>, but the small number of task conditions (<7) and the lack of a common measurement baseline have made it difficult to derive and evaluate task-based functional parcellations. The functional heterogeneity of the cerebellum has also been explored using meta analytic approaches<sup>9</sup>, which have the disadvantage that data for different tasks come from different groups of participants. In the present study, we

aimed to fully characterize the functional organization of the cerebellar cortex by employing a large and diverse task battery comprising 47 unique conditions designed to engage a broad range of sensorimotor, cognitive and social/affective processes. Using a block design, activation for each task was measured over four fMRI scanning sessions against a common baseline. Our task set was successful in eliciting activation across the entirety of the cerebellar cortex, allowing us to derive a novel parcellation that characterizes the functional profile of cerebellar subregions in unprecedented detail. The breadth of the task sets also enabled us to summarize the functional specialization of each region in terms of the underlying latent motor, cognitive and social/affective features. We developed a metric to evaluate the strength of the proposed functional boundaries. This allowed us to address the fundamental question of whether there are distinct functional regions in the cerebellum, or whether functional specialization is better described in terms of continuous gradients<sup>7</sup>. The approach is predicated on the idea that, if a boundary between two regions divides functionally heterogeneous regions, then the activation pattern for two voxels that lie within the same region should be more correlated than voxel pairs that span a boundary. Critically, a meaningful functional parcellation needs to be predictive of boundaries for the activation patterns elicited by a different set of tasks. Using this approach, we demonstrate that the cerebellum has discrete functional regions and that our multi-domain task battery (MDTB) parcellation is superior to alternatives in predicting functional boundaries. The new functional parcellation of the cerebellar cortex provides an important step toward understanding the role of the cerebellum across diverse functional domains.

## 1.3 Methods

### 1.3.1 Participants

All participants gave informed consent under an experimental protocol approved by the institutional review board at Western University. A total of 31 participants were scanned performing set A; 26 of this original cohort returned to perform set B (mean break between sessions = 166 d; s.d. = 153 d, with half returning about a year later and the other half having sessions separated by 2–3 weeks). The five participants who did not return for set B were not included in the analyses. Two additional participants were excluded from the analyses because they failed to complete all 32 scanning runs. Therefore, the final sample for the MDTB consisted of 24 healthy individuals (16 women, 8 men; mean age = 23.8 years, s.d. = 2.6) with no self-reported history of neurological or psychiatric illness. All participants self-reported as right-handed (Edinburgh Handedness Inventory > 40). The sample size was chosen to allow for an accurate assessment of inter-participant variability of the functional organization of the cerebellum.

### 1.3.2 Experimental tasks

The experimental tasks included in set A were chosen to engage a wide range of processing domains (cognitive, motor, affective, social), in many cases drawing on tasks

that had previously been shown to engage the cerebellum. While recognizing that our selection process was somewhat arbitrary and that the tasks would differ on a number of different dimensions, our main criterion was to use a large battery that broadly sampled different functional domains. A full description of the tasks, along with the accompanying references, is provided in Supplementary Table 1.

Set B included eight tasks that had been included in set A (shared tasks, for example, theory of mind, motor sequence) and nine unique tasks. The shared tasks provided a means to establish a common baseline across the two task sets. This enables between-task comparison across task sets, which is done by subtracting the mean activation pattern of the shared tasks from each task set. Only tasks that were successful at eliciting activation in the cerebellar cortex in set A were included as shared tasks in set B. For some of the novel tasks, we selected conditions that are thought to assay similar processing domains as in task set A. For example, both sets included working memory tasks, but the tasks involved different stimulus dimensions (for example, verbal working memory in set A and spatial mapping in set B). Other tasks (for example, the naturalistic movie-viewing tasks) were novel in task set B.

### 1.3.3 Experimental design

Each set consisted of 17 tasks. In every imaging run, each task was performed once for 35 s. The 35 s block was divided into a 5 s instruction period, where the task name (for example, 'theory of mind task'), the response effector ('Use your LEFT hand') and the button-to-response assignment ('1 = false belief, 2 = true belief') were presented on the screen. This was followed by a 30 s period of continuous task performance. In general, novel stimuli were introduced across imaging runs to prevent participants from learning specific stimulus–response associations. The one exception was the motor imagery task where participants were required to imagine playing a game of tennis. The number of trials within the 30 s block varied from 1 (for example, the movie-viewing and mentalizing tasks) to 30 (for example, go/no-go task). Most tasks involved 10–15 trials per block. The motivation for testing all tasks within a scanning run, as opposed to testing one task in each run, was to ensure a common baseline for all tasks, enabling between-task comparisons.

Three of the shared tasks (object 2-back, visual search, semantic retrieval) had a rapid, discrete trial structure (15 per block), whereby each unique stimulus (picture, letter, noun) was presented for 1.6 s, with the response required to be completed within this window, followed by an intertrial interval (ITI) of 400 ms. Three of the shared tasks had a slower discrete trial structure: motor sequence task (trials = 8; trial duration = 4.6 s; ITI = 400 ms); theory of mind (trials = 2; duration = 14.6 s; ITI = 400 ms); and action observation (trials = 2; duration = 14 s; ITI = 1 s). The remaining two shared tasks, spatial imagery and rest, did not have a discrete trial structure (duration = 30 s).

Of the nine unique tasks in set A, six had the rapid discrete trial structure (interval timing, International Affective Picture System (IAPS) affective, IAPS emotional, verbal

2-back, motor imagery, stroop, math, passive viewing: trial = 15, duration = 1.6 s, ITI = 400 ms; go/no-go: trials = 30; duration = 800 ms; ITI = 200 ms). The math task was comprised of 10 trials (duration = 2.6 s; ITI = 400 ms). The motor imagery task did not have a discrete trial structure (duration = 30 s).

Of the nine unique tasks in set B, six had a discrete trial structure: the prediction, spatial map, and response alternatives tasks entailed 6 trials per block (duration = 4.8 s; ITI = 200 ms), the mental rotation task 9 trials per block (duration = 3 s; ITI = 300 ms), the biological motion task 10 trials per block (duration = 3 s; ITI = 0 s) and the concrete permuted rules operations (CPRO) task 4 trials per block (duration = 7.3 s; ITI = 200 ms). The three movie-viewing tasks (landscape, animated and nature) did not have a discrete trial structure (duration = 30 s).

### 1.3.4 Hand assignment across tasks

For each task requiring responses, the responses were made with either the left, right or both hands using four-key button boxes. Hand assignment was consistent for sets A and B for the shared tasks. For tasks requiring two-choice discrimination, responses were made with the index or middle finger of the assigned hand, while responses for tasks requiring four-choice discrimination were made with the index and middle fingers of both hands. By including a motor feature model in our analysis, we could account for the motor requirements across the tasks.

### 1.3.5 Behavioral training

For each task set, participants completed 3 d of training before the first scanning session. Training included all of the tasks with the exception of the rest condition and the three movies (set B). For each set, the three training sessions took place over the course of 4–7 d (set A: mean number of days = 5.2, s.d. = 3.5; set B: mean number of days = 4.4, s.d. = 1.8).

The first day was used to familiarize participants with the requirements for each of the 17 tasks. Participants were instructed to carefully read the instructions. When ready, they initiated a 35 s training block. The number of training blocks differed depending on the perceived level of difficulty of the task. For example, the two-alternative forced choice picture-based tasks (IAPS affective, IAPS emotional) were practiced for three blocks, while the stroop task was practiced for seven blocks. During this training session, a run consisted of consecutive blocks of the same task. Online feedback was provided for response-dependent tasks (green or red squares to indicate correct or incorrect responses, respectively). At the end of each run, an overall accuracy score was provided concerning performance on the tasks requiring a button response.

On the second training day, switching between tasks was introduced. Participants were given 6 runs of training, with each run composed of 1 block for each of 11 tasks that required manual responses. As on day 1, the timing for the first four runs was

self-paced, with the participants allowed to read the instructions at their own pace before initiating the 30 s block. For the final two practice runs, the instruction phase was limited to 5 s, thus introducing the protocol that would be used in the scanner. Training on this day only included tasks that required overt responses. On the third training day, participants practiced all 17 tasks in four 10-minute runs (35 s per task), emulating the protocol to be used in the scanner sessions.

This training program ensured that participants were familiar with the requirements for each task and had considerable experience in switching between tasks. In this manner, we sought to minimize the impact of learning during the scanning sessions. On the third training day, performance was asymptotic, with participants correct on at least 85% of the trials for all of the tasks (range = 85–98%; see Supplementary Fig. 3).

### 1.3.6 Eye tracking

Eye tracking data were recorded on the third training session to obtain an estimate of saccadic eye movements for each of the tasks. An algorithm implemented in the Eyelink Toolbox (v. 1.5.0)<sup>31</sup> identified saccadic eye movements as events where eye velocity briefly exceeded a threshold of 30 degrees s<sup>-1</sup>. These data, tabulated as the mean number of eye movements per task, were included as a motor feature in the second-level feature model. Eye tracking data from two participants in set A and three participants in set B were not obtained due to technical problems. However, since eye movement behavior was consistent across participants, we used group-based estimates.

### 1.3.7 Scanning sessions

Participants completed four scanning sessions in total, two with set A and two with set B. The first scanning session for each set was conducted within a few days of the final training session (set A: mean = 2.0 d, s.d. = 1.6 d; set B: mean = 2.2 d, s.d. = 1.7 d) and the second scanning session was completed no more than 7 d after the first scanning session (set A: mean = 3.1 d, s.d. = 2.5 d; set B: mean = 2.7 d, s.d. = 2.3 d). Each scanning session consisted of eight imaging runs (10 min per run). Each of the 17 tasks was presented once for 35 s in each imaging run, producing 16 independent measurements per task. The task order was randomized across runs. To reduce order effects within each set, no two tasks were presented in the same order in two different runs. The order within each run, as well as the order of the runs, was kept constant for all participants. This procedure was chosen to allow for across-participant analyses on the time series level (results not presented in this article). As noted earlier, when possible, novel stimuli were used in each run to reduce the recall of specific stimulus–response associations.

### 1.3.8 Image acquisition

All fMRI data were acquired on a 3T Siemens Prisma at the Center for Functional and Metabolic Mapping at Western University. Whole-brain functional images were acquired

using an echo-planar imaging sequence with multi-band acceleration (factor 3, interleaved) and in-plane acceleration (factor 2), developed at the Center for Magnetic Resonance Research at the University of Minnesota. Imaging parameters were: repetition time = 1 s; field-of-view = 20.8 cm; phase encoding direction P to A; 48 slices; 3 mm thickness; in-plane resolution  $2.5 \times 2.5 \text{ mm}^2$ . Gradient echo field maps were acquired for distortion correction of the echo-planar imaging images due to B0 inhomogeneities (repetition time = 0.5 s, field-of-view = 24 cm, 46 slices with in-plane resolution of  $3 \times 3 \times 3 \text{ mm}^3$ ). We also acquired online physiological recordings of both heart and respiration during each functional run. For anatomical localization and normalization, a 5 min high-resolution scan of the whole brain was acquired (magnetization-prepared rapid acquisition gradient echo; field-of-view =  $15.6 \times 24 \times 24 \text{ cm}^3$ , at  $1 \times 1 \times 1 \text{ mm}^3$  voxel size).

### 1.3.9 Image preprocessing

Data preprocessing was carried out using tools from SPM12<sup>32</sup>, Caret (v. 5.65)<sup>33</sup> and spatially unbiased infratentorial template (SUIT v. 3.3)<sup>20</sup>, as well as custom-written scripts written in MATLAB 2015b. For all participants, the anatomical image was acquired in the first scanning session. Functional data were realigned for head motion within each session, and for different head positions across sessions using the six-parameter rigid body transformation. The mean functional image was then co-registered to the anatomical image and this transformation was applied to all functional images. No smoothing or group normalization was applied.

### 1.3.10 General linear model (GLM)

A GLM was fitted to the time series of each voxel separately for each imaging run. The 5 s instruction phase for all tasks was modeled using a single regressor, but was not included in later analyses. Each task was modeled using a boxcar regressor of 30 s, or a combination of multiple regressors if the block contained sub-conditions. These regressors could be 2 boxcar regressors of 15 s each (for example, the verb generation task where one sub-condition is word reading and the other is verb generation), 3 boxcar regressors of 10 s each (for example, visual search, display sizes of 4, 8 or 12) or 2 event-related regressors (for example, the stroop task, where each trial was congruent or incongruent). The rest condition was not modeled explicitly, but rather used as an implicit baseline in the model.

2.3.11 The quality of the GLM in modeling the blood-oxygen-level dependent signal response was determined by measuring the consistency of the activation patterns in the cerebellum across runs. This measure indicated that it was advantageous to omit the traditional high-pass filtering operation before the linear model (default operation in SPM). Instead, we opted to rely on the high-dimensional temporal autocorrelation model (the FAST option in SPM) to determine the optimal filtering, implemented in the GLM estimation. The beta weights from the first-level GLM were univariately pre-whitened by dividing them by the square root of the residual mean-square image. To include rest as a task condition in all subsequent analyses, we added a zero as an estimate for the rest

condition and then removed the mean for each voxel across all conditions. As such, the beta weights expressed the amount of activation elicited by each condition relative to the mean of all conditions.

To combine activation estimates across the two tasks sets, the mean of the shared tasks was removed separately for each set. Both sets were then combined, retaining the repeated estimates for the shared task. This resulted in a total of 61 estimates (set A = 29; set B = 32) for the 47 unique conditions. The activation patterns were re-centered by removing the overall mean across all 61 conditions.

### 1.3.11 Cerebellar spatial normalization

The SUI toolbox (v.3.3) in SPM12 was used to isolate the cerebellum from the rest of the brain and provide a normalization to a spatially unbiased template of the cerebellum<sup>34</sup>. The resulting cerebellar isolation mask was hand-corrected to ensure that it did not contain any shared voxels between the superior cerebellum and the directly abutting cerebral cortical regions of the inferior temporal and occipital cortex.

The probabilistic maps for the cerebellum were normalized into SUI space using the diffeomorphic anatomical registration through exponentiated lie algebra algorithm<sup>35</sup>. This algorithm deforms the cerebellum to simultaneously fit the probability maps of cerebellar gray and white matter onto the SUI atlas template. This non-linear deformation was applied to both anatomical and functional data. The activation estimates (that is, the beta weights) and residual mean-square images from the first-level GLM were resliced into SUI space. All images were masked with the cerebellar mask to avoid activation influences from the inferior occipital cortex. All data were visualized on a surface-based, flat-map representation of the cerebellar cortex in the SUI toolbox. The flat-map representation allows the spatial extent of task-evoked activation patterns to be fully visualized. Note that this flat map is not a true unfolding of the cerebellar cortex, but averages over a substantial number of folia. Therefore, it is meant for display purposes only<sup>14</sup>.

### 1.3.12 Motor feature model

Our primary goal was to study task-evoked activation patterns in the cerebellum beyond the well-known domain of motor function. Although not designed explicitly to measure motor-related activation, the 61 task conditions differed in the number of manual responses, as well as eye movements. To account for these motor-related activations, we generated a motor feature model (task conditions  $\times$  three motor features). For the hand movements, we entered the number of left- and right-hand presses for each task during the scanner runs. For eye movements, we used the group-averaged eye movement data from the third training day to estimate the number of saccades for each task condition. All motor features were encoded in terms of movements s-1 and z-normalized.

To extract and remove the motor-related activation across tasks, the 3 motor features were combined with an indicator matrix that had a 1 for each of the 61 task conditions. To estimate and subsequently remove the influence of the motor features, we estimated the linear model with L2-norm regression (fixed  $\lambda$  of 0.01) from the beta estimates of each participant (task conditions  $\times$  voxels  $\times$  participant). The average of all task conditions was used as a baseline measure and subtracted from the motor-corrected activation estimates. The activation estimates for the shared tasks were first averaged; then a group average was computed for the purposes of visualization on the cerebellar flat map<sup>14</sup>.

### 1.3.13 Reliability of activation patterns

To determine intra-participant reliability across the entire cerebellar cortex, we calculated the correlation between the average activation estimates for the first and second session for each task set, separately for each participant. To obtain an overall reliability, we stacked the 29 (A) or 32 (B) activation estimates for all cerebellar voxels into a single vector and calculated the Pearson correlation between the two estimates. For Fig. 1e, this analysis was also performed for each voxel separately. The group-averaged correlations were then visualized on the cerebellar flat map.

### 1.3.14 Spatial frequency of activation patterns

To determine how much of the variance of the activation patterns was common to the group relative to how much was idiosyncratic to the individual participants, we calculated two correlations, one between task-activity maps between two sessions for the same participant (as for reliability) and the second between sessions of different participants. Correlations were computed on all gray matter voxels in SUIT space. To determine the spatial scale of these common activation patterns, we decomposed the volume image for each task condition into five spatial frequency bands ranging from 0 to 5 cycles  $\text{cm}^{-2}$ . This decomposition was done separately for each participant, study, session and task condition. The within- and between-participant correlations were then computed for each spatial frequency band.

### 1.3.15 Evaluating functional boundaries

We developed a novel method to evaluate functional boundaries from fMRI data. The rationale of the method is that, if a boundary is dividing two functionally heterogeneous regions, then two voxels that lie within the same region should have more similar functional profiles than two voxels that are in different regions (Fig. 2a; Eq. 1). Because functional organization tends to be smooth, the correlation between two voxels will be higher for two adjacent voxels, and fall off as the spatial distance increases<sup>11</sup>. To control for distance, we calculated the activation pattern correlations for all pairs of voxels separated by a fixed Euclidean distance, using spatial bins ranging from 4 mm to 35 mm. To exclude spatial correlations that were driven by noise, we used a cross-validated correlation. Using  $\mathbf{u}_{i,1}$  to represent the functional profile (zero-meaned) of voxel  $i$  from one session, and  $\mathbf{u}_{j,2}$  the functional profile of voxel  $j$  from the other session, the correlation was calculated as:

$$r = \frac{\frac{1}{2} \sum_{i,j} (u_{i,1}^T u_{j,2} + u_{i,2}^T u_{j,1})}{\sqrt{\sum_{i,j} (u_{i,1}^T u_{i,2}) \sum_{i,j} (u_{j,1}^T u_{j,2})}},$$

Equation 1

where the sum was done on all voxel pairs  $i,j$  in the corresponding 5 mm bin, separating those that involved voxel pairs from the same region (within) and those in which the voxel pairs spanned the boundary (between) for each spatial bin. We excluded voxels where the term  $u_{i,1}^T u_{i,2}$  was negative, as it indicated the absence of any reliable tuning across the two sessions. The difference between the within-region correlation and the between-region correlation defined the Distance Controlled Boundary Coefficient (DCBC). A positive DCBC value indicates that voxel pairs originating from the same region are more functionally related than voxel pairs that lie across boundaries. The DCBC was calculated for each participant and spatial bin separately, and then averaged.

The DCBC can serve not only as a global measure of a parcellation (averaging across the cerebellum and spatial bins), but also as a measure to evaluate the strength of individual boundaries. For the latter, we first identified boundaries using an edge-based connectivity scheme<sup>37</sup>. The strength of a given boundary is defined by the DCBC calculated only on the voxel pairs from the two regions that are separated by that boundary. To visualize boundary strength, the thickness of the boundary on the flat map was based on its DCBC value.

We applied this boundary evaluation procedure to MDTB parcellations, as well as parcellations based on lobular boundaries or task-free fMRI data. The lobular parcellation was obtained from a probabilistic atlas of the human cerebellum<sup>20</sup> that includes regions for lobules I–IV, V, VI, Crus I, Crus II, VIIb, VIIIa, VIIIa, IX and X. To ensure that the poor performance of the lobular parcellation was not due to inaccuracies in detecting the lobular boundaries, we repeated the analysis using a manual lobular parcellation in five participants. The parcellation from this sample predicted functional boundaries about as well as the one derived from the probabilistic atlas (DCBC: 0.022 versus 0.025;  $t_5 = 1.441$ ,  $P = 0.209$ ). The task-free 10-region parcellation<sup>6</sup> was based on data archived as part of the HCP, while the other two were based on a large 1,000-person data set collected at Harvard and Massachusetts General Hospital<sup>5</sup>. All parcellations were sampled into SUIT space and evaluated using our MDTB.

### 1.3.16 MDTB data set parcellation

To derive a parcellation from the MDTB, we used the activation profiles of gray-matter voxels averaged in SUIT volumetric space across participants. We used convex non-negative matrix factorization<sup>38</sup> to decompose the  $N$  (tasks)  $\times$   $P$  (voxels) data matrix into a product of an  $N \times Q$  (regions) matrix of task profiles and a  $Q \times P$  matrix of voxel weights. The voxel weights, but not the task profiles, are constrained to be non-negative. Furthermore, the task profiles are convex combinations of the raw data. Compared to other decomposition methods, such as independent component analysis,

this method has the advantage that voxels cannot be explained by an inverted or negative regional task profile. This constraint is also reasonable given that mossy fiber input, the main neural signal driving the blood-oxygen-level dependent signal response in the cerebellum, is excitatory<sup>25</sup>. To ensure convergence, we started the decomposition with random initializations and selected the iteration with the best reconstruction of the original data. A winner-takes-all approach was adopted to assign each voxel to the region with the highest weight.

To allow for a direct comparison with existing task-free parcellations<sup>5,6</sup>, we used parcellations with 7, 10 and 17 regions (Supplementary Fig. <sup>5d-f</sup>). Parcellations involving regions within this range achieved similar reconstruction accuracy and quality of functional boundaries.

We also derived separate parcellations for each participant to determine whether boundaries could be predicted better using an individual approach. An advantage of the individual parcellation is that idiosyncrasies of within-participant organization are captured. The disadvantage is that individual parcellations are derived on substantially fewer data than the group.

### 1.3.17 Bootstrap analysis

To obtain a measure of boundary uncertainty, we performed bootstrap analyses across both participants and task conditions. For participants, we repeatedly drew 24 participants (with replacement) from our sample, averaged the data and derived a new functional parcellation. To be able to relate the parcellations to each other, each parcellation used the original solution as a starting value. For the task conditions, we repeatedly drew 47 task conditions (with replacement) from our data, again deriving a new parcellation each time. For each analysis, we repeated this process 100 times.

To evaluate the consistency of the parcellations globally, we calculated the adjusted Rand index, which measures the correspondence between two parcellations (0 = overlap not different from chance, 1 = perfect overlap). For a regionally specific analysis, we counted the number of times that each voxel was assigned to the same (most frequent) region. For visual display (Fig. 3d), we then used this assignment certainty to determine the transparency of the region coloring (<50% = fully transparent, 100% = fully opaque).

### 1.3.18 Evaluation of functional parcellations

To evaluate the group and individual MDTB parcellations, we wanted to know how well functional boundaries could be predicted for each participant using a completely novel set of tasks. Because we did not acquire data with a third, independent task set, we used the existing data to estimate lower and upper bounds of predictability. For the lower bound, we derived the parcellation using the data from set A and evaluated it with the data from set B, using the unique tasks only. This procedure was then repeated with

the task sets reversed and the results were averaged across the two cross-validation folds. Note that the outcome of this analysis will probably result in a lower value than would be obtained with the final segmentation, since each parcellation is based on half of the available data. As such, we used this estimate as an approximate lower bound. We also evaluated a parcellation derived from both sets A and B. We evaluated this parcellation as before, excluding the shared tasks from both task sets to make the estimate consistent with the lower-bound estimate. Because there is overlap between the data used for training and evaluation, the performance measure is overfitted and, therefore, was taken as an approximate upper bound. The true performance of the full parcellation, if applied to a completely new task set, would probably fall between these lower and upper bounds.

The anatomical and task-free parcellations could be directly evaluated on the MDTB data since each parcellation was derived from independent data. For consistency, we excluded the data from the shared tasks in the evaluation set.

To evaluate the degree to which the results depended on the similarity of the tasks in the training and test sets, we repeated the analysis, this time selecting the seven most distinct task conditions in each test set (Supplementary Fig. 6). The conditions were selected by computing the distance between activity patterns for each test condition to each training condition (Supplementary Fig. 4a). Then, we identified for each test condition the closest match in the training set and selected the seven test conditions for which this closest match was most dissimilar.

To further validate our results, we evaluated the MDTB and task-free parcellations on the task-based data from the HCP data set (<https://db.humanconnectome.org>)<sup>16</sup>. We used data from the 214 most recently added participants (scanned at 3T). Of the 214, 186 participants had complete data sets and these constituted our final sample. For each participant, we evaluated the parcellations on a set of 22 contrast maps from 7 tasks (all against rest).

### 1.3.19 Representational structure of task-related activation patterns

Representational similarity analysis<sup>17</sup> was used to investigate the representational structure of task-related activation patterns from the MDTB cerebellar data. The dissimilarity between the motor-corrected activation patterns was measured for each pair of task conditions using the cross-validated Mahalanobis distance, using the imaging runs as independent partitions<sup>39</sup>. To calculate the distances between conditions across the sets, we first subtracted the mean of the shared task conditions from each imaging run. Cross-validation ensures that the average (expected) value of the dissimilarity measure is 0 if the two activation patterns only differ by noise. This allowed us to test for significant differences between activation patterns using a one-sample *t*-test against 0.

Classical multidimensional scaling was employed to visualize the distances between all possible pairs of task conditions. For the purposes of visualization, the pairwise distances for the shared tasks were averaged so that there were 47 (rather than 61) task conditions in the representational dissimilarity matrix. Multidimensional scaling projects the  $N$ -dimensional representational dissimilarity matrix into a lower-dimensional space so that distances from the higher space are preserved with as much integrity as possible. Multidimensional scaling was performed on the group-averaged representational dissimilarity matrix and the first three dimensions were visualized in a three-dimensional space.

### 1.3.20 Feature-based approach

The power of our task-based approach in studying the cerebellum is that we can identify the involvement of each region across functional domains and different task variations. To summarize the task activation profiles for each region, we used a feature-based encoding model. The features included the three motor features (see earlier) and cognitive features, selected to capture the hypothetical mental processes involved in each task. To derive these features, we used an online cognitive ontology<sup>18</sup>, an atlas of tasks and the concepts associated with those tasks. Of the 815 concepts currently included in the atlas, 46 were judged to provide an appropriate and sufficient characterization of the tasks in our battery, creating a feature matrix (47 task conditions  $\times$  46 features). For example, features such as semantic knowledge and lexical processing were associated with tasks such as verb generation and semantic prediction; emotion recognition was associated with the IAPS emotional processing task and the biological motion task. As with the motor feature model, each feature was z-standardized and feature weights for each region were estimated with non-negative regression. For visualization purposes, the three highest weights for each region were computed.

### 1.3.21 Statistical analysis and inference

Unless otherwise noted, all statistical tests were based on the  $n = 24$  participants of our sample, considering participant as a random effect. Therefore, all  $t$ -tests were repeated measures with 23 d.f. Only the analysis presented in Fig. 4e is based on the HCP task data and was therefore based on  $n = 186$  participants. All  $t$ -tests were two-sided. Data distribution was assumed to be normal, but this was not formally tested.

## 1.4 Results

To obtain a comprehensive functional parcellation of the cerebellar cortex, we developed an MDTB of 26 tasks comprising 47 unique task conditions (Fig. 1a and Supplementary Table 1), selected to encompass a wide range of processes required for motor, cognitive and affective/social function. To avoid strong learning-related changes, 24 healthy individuals were trained on the task protocol (approximately 10h) before scanning. During scanning, each task was performed once per imaging run for a 35 s block (Fig. 1b). This ensured that all tasks were measured against a common baseline,

allowing for any between-task comparison. To make this approach feasible, the tasks were split into two sets (Fig. 1a), and each task set was tested in two separate fMRI scanning sessions, resulting in a total of approximately 5.5h of functional data per participant.

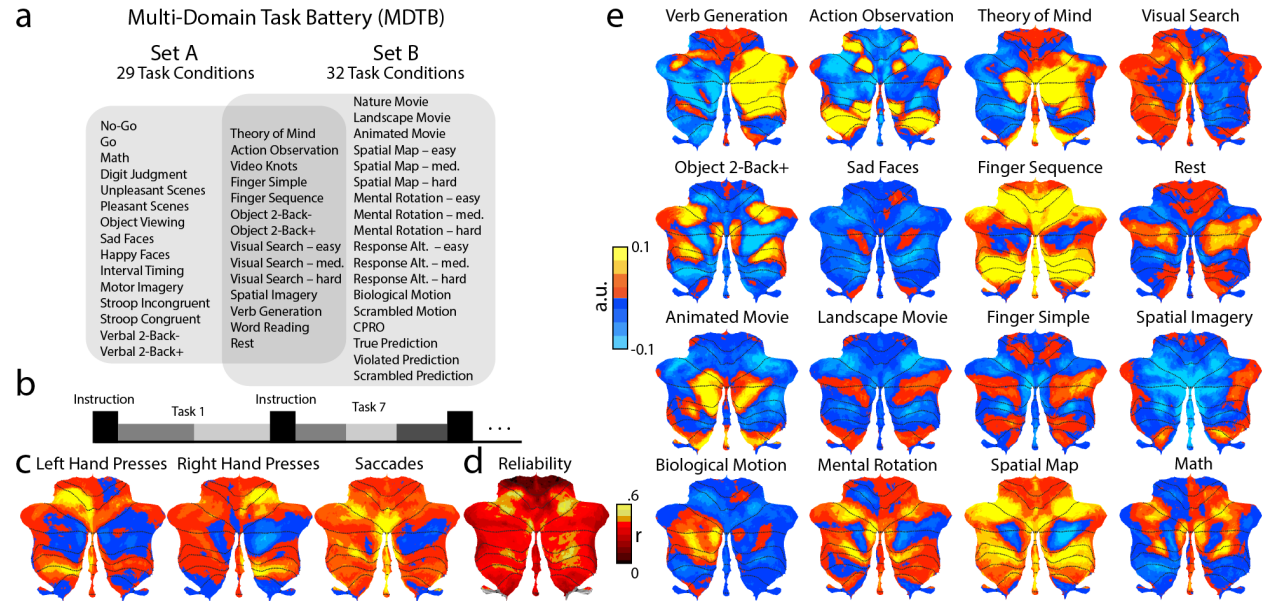


**Fig 1. Multi-Domain Task Battery (MDTB).** *a*, Experimental design. A total of four fMRI scanning sessions were collected on the same set of participants using two tasks sets. Each set consisted of 17 tasks, with 8 tasks in common. The tasks were modeled as 29 task conditions in set A and as 32 in set B, with 14 task conditions common across both task sets. *b*, Timing of each task: 5 s instruction period followed by 30 s of task execution. Tasks consisted of a different number of task conditions (gray bars, range 1–3). *c*, Unthresholded, group-averaged motor feature maps, displayed on a surface-based representation of the cerebellar cortex<sup>14</sup>. *d*, Across-session reliability of activation patterns for each voxel. *e*, Group-averaged activation maps for selected tasks, corrected for motor features. The red-to-yellow colors indicate increases and the blue colors denote decreases in activation, relative to the mean activation across all conditions. Activity is normalized by the root-mean-square error of the time series fit for each voxel. CPRO, concrete permuted rules operations.

#### 1.4.1 Identification of motor features from the MDTB.

As a first step, we sought to identify cerebellar regions where the hemodynamic response was closely tied to motor function, specifically hand and eye movements. Our experimental design did not include specific contrasts that isolated each motor component. Instead, we varied the motor demands across task conditions; for example, the motor sequence task involved approximately 40 left and right finger responses, the theory of mind tasks two left-hand responses and the movie tasks no responses. We then generated a motor feature model, which included the number of left- and right-hand responses and the number of saccadic eye movements made per task (see Methods). Using regularized regression, we could then estimate the activation across

tasks attributable to motor involvement. Left- and right-hand movements were associated with activation increases in the two hand motor areas of the cerebellum (Fig. 1c), the anterior hand region located on the boundary of lobules V and VI, and the inferior region in lobules VIIIb<sup>10</sup>. Saccadic eye movements elicited activation in vermis VI, consistent with the location of the oculomotor vermis in the macaque monkey<sup>11</sup>. Compared to previous contrast-based human fMRI studies<sup>12</sup>, which have yielded relatively inconsistent results, our feature-based mapping approach resulted in an extraordinarily clear localization of eye movement activation to the oculomotor vermis. While these results mainly confirm the well-known functional localization within the cerebellum for movement, they demonstrate that a broad task-based approach without tightly matched control conditions provides a powerful means of revealing functional organization.

#### 1.4.2 MDTB elicits varied activation patterns across the cerebellum.

We then characterized the task-related activation patterns that could not be explained by basic motor features. Overall, we could elicit strong and distinguishable patterns of activation (Fig. 1e and Supplementary Fig. 1) across the cerebellar cortex. To determine the reliability of the activation patterns, we calculated the correlation of the individual, unsmoothed task activation profiles for each voxel across the two sessions of each set. On average, these task activation profiles were reliable (set A:  $r=0.43$ , 95% confidence interval (CI)=0.39–0.46; set B:  $r=0.42$ , 95% CI=0.37–0.46; see Supplementary Fig. 2 for individual participants). The resulting voxel-wise reliability map (Fig. 1d) confirmed that this was the case for the entire cerebellar cortex, with the exception of lobules I–IV. These lobules are associated with foot movements<sup>10,13</sup>, a feature absent from our tasks. Qualitatively, the activation patterns elicited by our task sets replicated numerous results obtained in previous neuroimaging studies. For example, right-lateralized activation throughout Crus I, Crus II and VIIb was observed with the verb generation task<sup>8</sup> while left-lateralized activation throughout Crus I and Crus II was demonstrated with the biological motion task. Consistent with previous working memory studies<sup>14</sup>, the n-back tasks activated two distinct lateral regions of lobules VII. Recent evidence for medial Crus I and Crus II activation during movie tasks was also corroborated<sup>15</sup>. The task activation maps also demonstrated some insights that have not been (or not as clearly) reported in the previous literature. The rest condition (contrasted against the mean of all the other conditions) was associated with bilateral activation in a mid-hemispheric region in Crus I and II, effectively forming the cerebellar component of the default mode network<sup>5</sup>. Similar cerebellar regions were strongly activated during the theory of mind<sup>14</sup> and movie tasks<sup>15</sup>. The finger sequence and visual search tasks led to strong activation in cerebellar hand and eye movement-related areas, respectively. Given that these activation maps were corrected for movement-related activity, these results indicate that these areas are especially activated during movements with high attentional demands. Finally, the action observation task elicited activation in a distinct set of areas surrounding the motor areas of the cerebellum, especially in the posterior motor representation. When using a dissimilarity measure to construct a representational space for all tasks (Supplementary Fig. 4), the action observation condition emerged as one of the most unique activity patterns. The passive picture viewing tasks (that is, sad faces) did not elicit much activation in the cerebellum. This is generally consistent with

the notion that the cerebellum does not receive cortico-pontine projections from the inferior temporal cortex, a pathway involved in visual object and scene recognition. To quantify this observation, we tested the activation patterns of all possible task conditions against each other. While over 95% of the pairwise comparisons were significant (uncorrected  $p < 0.001$ ), the most notable exceptions were pairs of the picture viewing tasks (Supplementary Fig. 4a). In contrast, passively watching engaging movie snippets (nature movie, animated movie) resulted in reliable and specific activity patterns (Fig. 1e and Supplementary Fig. 1), probably related to processes required for action perception and social cognition.

### 1.4.3 Cerebellar lobules do not reflect functional subdivisions.

One way to summarize these activation patterns is to subdivide the cerebellum into functionally distinct regions. However, this approach is only meaningful if there are stable functional subdivisions in the cerebellum that generalize across tasks. To address this fundamental question, we developed an evaluation metric, which we refer to as the distance-controlled boundary coefficient (DCBC). If a boundary divides two functionally heterogeneous regions, then any equidistant pair of voxels within a region should have activation profiles that are more correlated with each other than two voxels that are separated by the boundary (Fig. 2a; see Methods). Specifically, we calculated correlations between voxel pairs using a range of spatial bins (4–35mm). The difference between the within- and between region correlations for each spatial bin then served as our evaluation criterion. This method extends standard clustering metrics (that is, silhouette coefficient) to account for spatial distance. Given that the spatial resolution of fMRI is insufficient to cleanly resolve individual folia, the spatial distance was measured in the volume (see Methods for details).

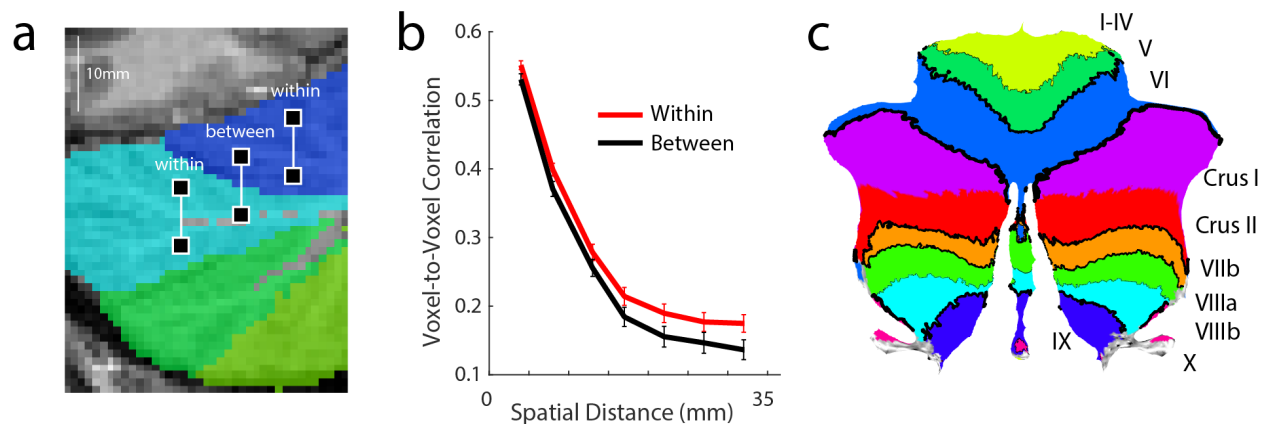


**Fig 2. Distance-Controlled Boundary Coefficient (DCBC).** *a*, Correlations between all pairs of voxels with the same distance were calculated and averaged depending on whether they were ‘within’ or ‘between’ regions. Voxel pairs were then binned according to spatial distance in the volume (4–35 mm in steps of 5 mm). *b*, Average cross-validated correlation (see Methods) as a function of spatial distance for lobular boundaries. The DCBC is defined as the difference in correlation (within-between) within each distance bin. The error bars show between-participant s.e.m. for  $n = 24$

*participants. c, Strength of the boundaries for lobular parcellation, with the thickness of the black lines indicating the DCBC value.*

We first employed this evaluation method to determine the degree to which functional boundaries follow the major lobular subdivisions<sup>5</sup>. This is a question of high practical importance given that lobular boundaries are commonly used to define regions of interest for interpreting functional activations in the cerebellum. Notably, the correlation between voxels within a lobule was not much greater than the correlation between voxels that spanned a lobular boundary (Fig. 2b). The correlations, averaged over distances of 4–35mm, were  $r=0.28$  (95% CI, 0.26–0.30) within lobules and  $r=0.25$  (95% CI, 0.05–0.46) between lobules. While statistically significant ( $t_{23}=4.62$ ,  $p < .01$ ), the difference was very small (DCBC=0.03). Thus, lobular boundaries do not reflect strong functional subdivisions in the cerebellum. The DCBC can also be used to evaluate the strength of individual boundaries (Fig. 2c). For example, the superior posterior fissure separating lobule VI from VII was the strongest lobular boundary (DCBC=0.152), while the primary fissure, which serves as the first principal subdivision of the cerebellum, was relatively weak (DCBC=0.068). The boundary separating Crus I and Crus II did not predict any functional specialization (DCBC=0). In sum, many cerebellar fissures did not demarcate a change in function.

#### 1.4.4 MDTB parcellation uncovers strong functional boundaries.

Next, we asked whether a parcellation based on the MDTB data would more clearly identify functional boundaries. We first estimated a group-based parcellation using all of the MDTB data. Using convex semi-non-negative vector factorization, we decomposed the  $N$  (tasks)  $\times$   $P$  (voxels) data matrix into a product of an  $N \times Q$  (regions) matrix of task profiles and a  $Q \times P$  matrix of voxel weights. The voxel weights, but not the task profiles, were constrained to be non-negative. Using a winner-takes-all approach, we then assigned each voxel to the region with the highest weight. Figure 3b shows the resulting parcellation using ten regions. For this parcellation, the average DCBC was 0.159 (Fig. 3a, dashed line,  $t_{23}=31.85$ ,  $p < 1 \times 10^{-10}$ ), a value higher than that obtained for the strongest lobular boundary.

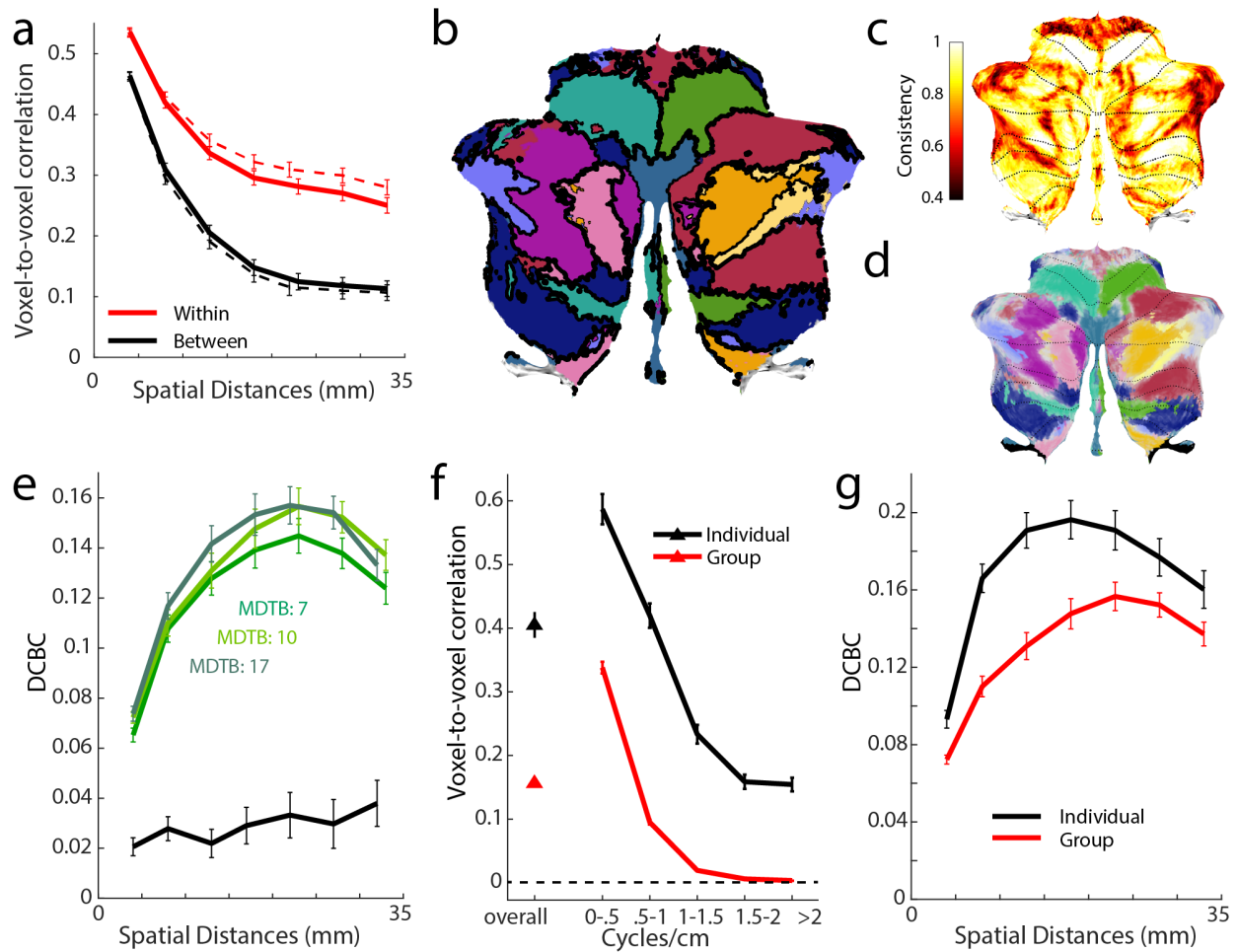


**Fig. 3: MDTB parcellation reveals functional boundaries in the cerebellar cortex.** *a*, Average cross-validated correlations (see Methods) for ‘within’ (red) and ‘between’ (black) voxel pairs for MDTB parcellation (10 regions). The solid lines indicate the values for the cross-validated estimates; the dashed lines are the estimates for the full parcellation. *b*, Ten-region MDTB parcellation (see Supplementary Figure 8 for individual participant parcellations). The DCBC for each boundary is visualized by the thickness of the black lines. *c*, Proportion of samples in the bootstrapped analysis (participants) where the voxel was assigned to the same compartment as in the original parcellation. Most voxels had a consistency of assignment  $>0.6$ . *d*, Visualization of boundary uncertainty, using the color scheme in *b*, but adjusted so that the degree of transparency is indicative of the uncertainty of the assignment. Voxels that were assigned to a single compartment on less than 50% of the cases are shown in gray. *e*, DCBC as a function of the spatial distance for the lower bound of the three MDTB parcellations (colored lines) and lobular parcellation (black line). *f*, Within-subject (black) and between-subject (red) reliability of activation patterns overall and across different spatial frequencies. *g*, DCBC as a function of voxel distance for the ten-region group parcellation (red) and the average of the ten-region individual parcellations (black). Only the cross-validated estimates of the prediction performance for novel tasks are shown. In all panels, the error bars show the between-participant s.e.m. for  $n = 24$  participants.

However, functional parcellations will invariably yield boundaries for a given task set, since training and evaluation data overlap. Critically, a good parcellation should be able to predict boundaries for a new set of tasks. Therefore, we determined the parcellation based on all task conditions from set A and evaluated the boundaries using the unique tasks from set B. We repeated this out-of-sample generalization test in the other direction and averaged the two values. Using this approach, the average DCBC was 0.130, only slightly lower than the non-cross-validated estimate (Fig. 3a;  $t_{23}=24.232$ ,  $p < 1 \times 10^{-10}$ ). The cross-validated DCBC will underestimate the true predictive power of the full parcellation, with true performance on a novel task probably falling between the cross-validated and non-cross-validated DCBC. To remain conservative, we only report the cross-validated DCBC estimates for the remainder of the article. The exact form of a parcellation depends on the specified number of regions. We also derived a parcellation with 7 (Supplementary Fig. 5d) or 17 regions (Supplementary Fig. 5f). While the 7-region parcellation performed slightly worse than the 10-region parcellation (DCBC=0.121,  $t_{23}=-4.18$ ,  $p = 0.00036$ ), the 17- and 10-region parcellations performed comparably (DCBC=0.133,  $t_{23}=1.57$ ,  $p = 0.131$ ; Fig. 3e). While there was reasonable agreement across the different MDTB parcellations (Supplementary Fig. 5h), some differences in the functional subdivisions for the different parcellations emerged. While our results clearly show that the MDTB parcellations reflect true functional boundaries in the cerebellum, they also make clear that there are a number of equivalent ways to subdivide the cerebellum. Thus, the exact choice of a ‘final’ parcellation is constrained by practical considerations. In this study, we focus on the ten-region parcellation because it provides a useful level of resolution for a full functional characterization. To assess the stability of the parcellation, we conducted a bootstrap analysis, across both participants and task conditions (for the details, see Methods). The mean Rand coefficient between each of the new parcellations and the original parcellation was 0.646 (95% CI=0.55–0.73) for the bootstrap across participants and 0.654 (95% CI=0.58–0.73) across task conditions. To quantify the uncertainty of specific boundaries, we calculated the proportion of bootstrap samples for which each voxel was assigned to the same compartment as in the original parcellation (Fig. 3c). Overall, consistency was good for most of the cerebellum (Fig. 3d). Lobules I–IV had higher uncertainty, probably a consequence of a lack of foot movements in our task battery. The parcellations described earlier were based on group data. To quantify the variability in functional organization across individuals, we compared the correlation between the task activation maps across participants to the within-participant reliability across the two sessions (Fig. 3f). Overall, 27.7% of the pattern variance was shared between individuals, whereas 72.3% reflected idiosyncratic patterns. A spatial frequency decomposition of the patterns (see Methods) revealed that commonalities across participants were restricted to the low spatial frequencies ( $<1$  cycle  $\text{cm}^{-2}$ ; activations of more than 5mm in size), while the fine-grained patterns were purely idiosyncratic to the participant. Indeed, a parcellation derived from the functional data from the individual significantly outperformed group parcellation in predicting functional boundaries for new tasks for that same individual (Fig. 3g;  $t_{23}=5.88$ ,  $p < 1 \times 10^{-5}$ ). In summary, using the MDTB data, we were able for the first time to quantitatively demonstrate the existence of distinct functional regions in the human cerebellum. Our results clearly advocate the

adoption of a functional parcellation to replace lobular subdivisions as a tool to summarize functional cerebellar data.

#### 1.4.5 Task-free parcellations identify overlapping but weaker boundaries.

Prior work has leveraged the correlational structure of taskfree (or 'resting state') fMRI data to derive various parcellations of the cerebellum, using 75 , 106 or 175 regions (Fig. 4a–c). These parcellations were only moderately consistent with each other (Fig. 4f), with an average adjusted Rand index of 0.33 (0=no communality; 1=perfect match). Correspondence between the different MDTB parcellations was slightly higher (adjusted Rand index=0.47), indicating more stability across the MDTB parcellations. The average adjusted Rand index between the MDTB and task-free parcellations was 0.15, indicating that there are systematic differences between the two approaches. To determine where task-free and MDTB parcellations diverge, we conducted a searchlight analysis, computing the adjusted Rand index locally using a 1-cm radius sphere for each pair of parcellations. The results demonstrated that task-free and MDTB parcellations corresponded most tightly in the mid-lateral areas of lobule VII. In these 'default mode' regions, the agreement between the MDTB and task-free maps (Fig. 4g) was similar to the agreement between MDTB maps (Supplementary Fig. 5h). In contrast, in more lateral aspects of lobule VII, and especially areas engaged in motor control or action observation, the correspondence between task-free and MDTB parcellations was much weaker. This is probably due to the relatively low consistency among the task-free parcellations (Supplementary Fig. 5g).

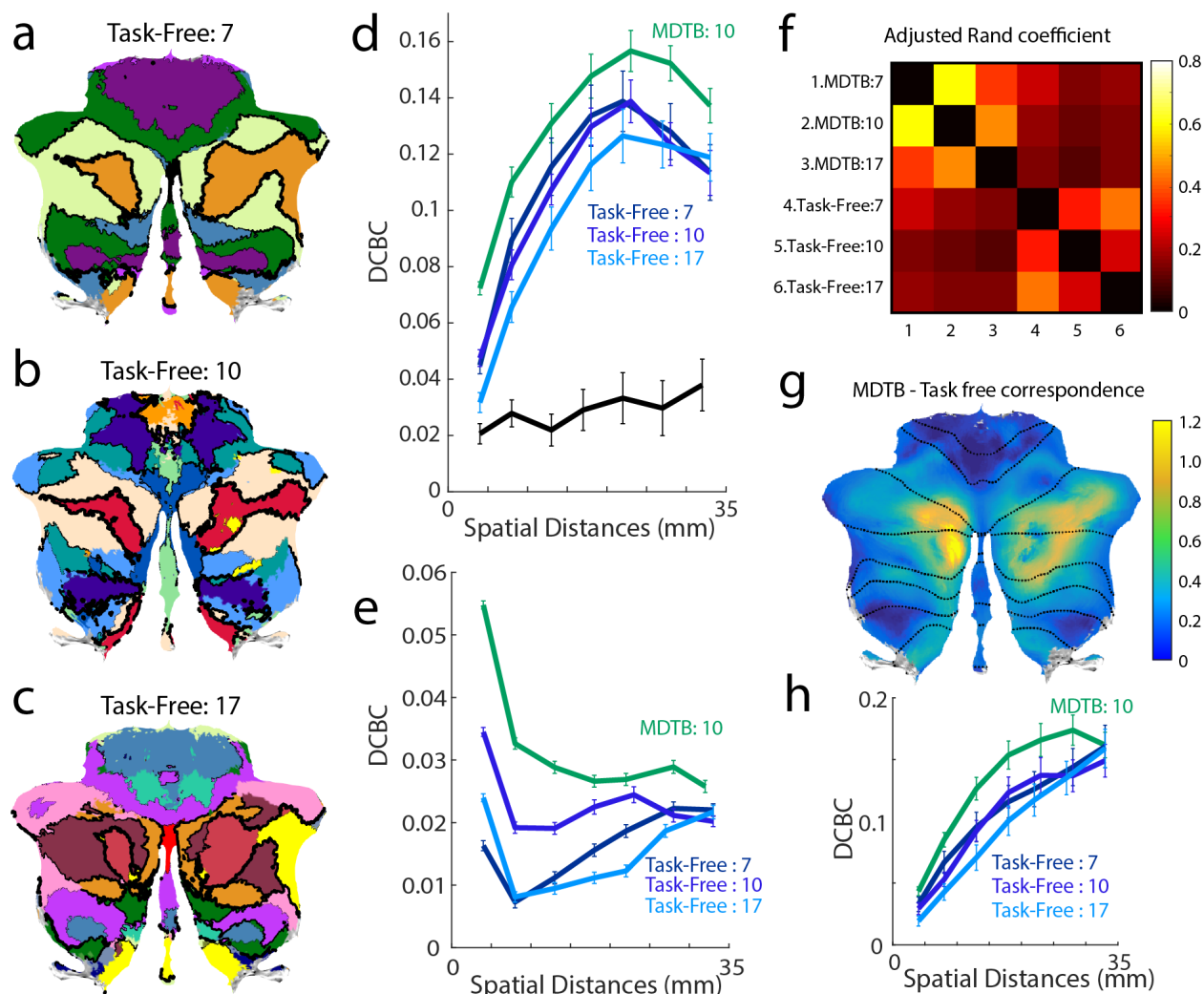


**Fig. 4: Comparison of MDTB and task-free parcellations of the cerebellum.** a–c, Cerebellar parcellations based on task-free data with 7, 10 and 17 regions<sup>6</sup>. Thickness of the black lines indicates the DCBC for the corresponding boundary. d, DCBC for different spatial distances for lobular (black), task-free (dark and light blue) and MDTB (green) parcellations. MDTB parcellation was evaluated in a cross-validated fashion (see text). e, Evaluation of the same parcellations on task-based data from 186 HCP participants. f, Matrix of adjusted Rand coefficients between three versions of the MDTB parcellations (7, 10, 17) and the three task-free parcellations (7, 10, 17). g, Correspondence between MDTB and task-free parcellations. A value of 1 (yellow-green) indicates that the adjusted Rand coefficient between task-free and MDTB parcellations is the same size as the adjusted Rand coefficient between MDTB parcellations. Lower values (blue-green) indicate weaker agreement between task-free and MDTB parcellations compared to MDTB on its own. h, Evaluation of MDTB parcellation (derived only from set A) and the task-free parcellations on the three movie tasks from set B. d, e, h, The error bars show the between-participant s.e.m. for  $n = 24$  or  $n = 186$  (e) participants.

We then evaluated whether task-free parcellations could predict functional boundaries in our MDTB data. The average DCBC for the task-free 7-, 10- and 17-region parcellations was 0.109, 0.106 and 0.097 respectively, substantially higher than the lobular parcellation ( $t_{23}=16.849$ ,  $p < 1 \times 10^{-10}$ ; Fig. 4d). Thus, all of the task-free parcellations are, to some degree, able to predict functional boundaries. However, the average task-free DCBC was significantly lower than the ‘lower bound’ for our MDTB 10-region parcellation ( $t_{23}=5.585$ ,  $p < 1 \times 10^{-5}$ ). This indicates that the MDTB parcellation outperformed the task-free parcellations in predicting functional boundaries on a novel set of task conditions. Although the task conditions used for evaluation did not overlap with the tasks used for deriving the MDTB parcellation (see also Supplementary Fig. 6), we wanted to ensure that the superior performance of the MDTB parcellation would generalize to a completely separate data set. To this end, we evaluated the MDTB and task-free parcellations using data from 186 participants from the task-based Human Connectome Project (HCP; Fig. 4e) (ref. 16). Again, the MDTB 10-region parcellation significantly outperformed the three task-free parcellations (7-region:  $t_{185}=22.671$ ,  $p < 1 \times 10^{-10}$ ; 10-region:  $t_{185}=13.266$ ,  $p < 1 \times 10^{-10}$ ; 17-region:  $t_{185}=28.09$ ,  $p < 1 \times 10^{-10}$ ). To ensure that the higher predictive power of the MDTB parcellation was not solely driven by regions associated with motor control, we reevaluated the DCBC using only the three movie tasks (Fig. 4h). Even though these conditions did not demand any overt movement, the advantage of the MDTB over the 7-region ( $t_{23}=2.7$ ,  $p = 0.01$ ), 10-region ( $t_{23}=5.3$ ,  $p < 1 \times 10^{-5}$ ) and 17-region ( $t_{23}=5.1$ ,  $p < 1 \times 10^{-5}$ ) task-free parcellation remained significant. Overall, these results demonstrate that the advantage of the MDTB over task-free parcellations extends to new data sets and to conditions that do not involve active tasks.

#### 1.4.6 Characterizing activation by cognitive features.

An important advantage of a task-based approach is that we can make inferences about the processes that activate the cerebellar cortex. To characterize the functional profiles in each of the regions across tasks, we used predefined and non-orthogonal features<sup>17</sup>. We already successfully applied this approach when characterizing the activation patterns elicited by motor features (Fig. 1c), which could be directly operationalized as the number of finger and eye movements. To extend this approach, we needed to describe each task condition in terms of its underlying cognitive features. Therefore, we turned to the Cognitive Atlas, an online cognitive ontology<sup>18</sup>, which summarizes the current consensus in cognitive science of the processes associated with a large array of tasks. To construct a feature space, each of the task conditions was rated on each of the cognitive concepts (see Methods). We then estimated the feature weights for each region using non-negative regression. For visualization purposes, we depicted the top three feature weights for each region (Fig. 5).

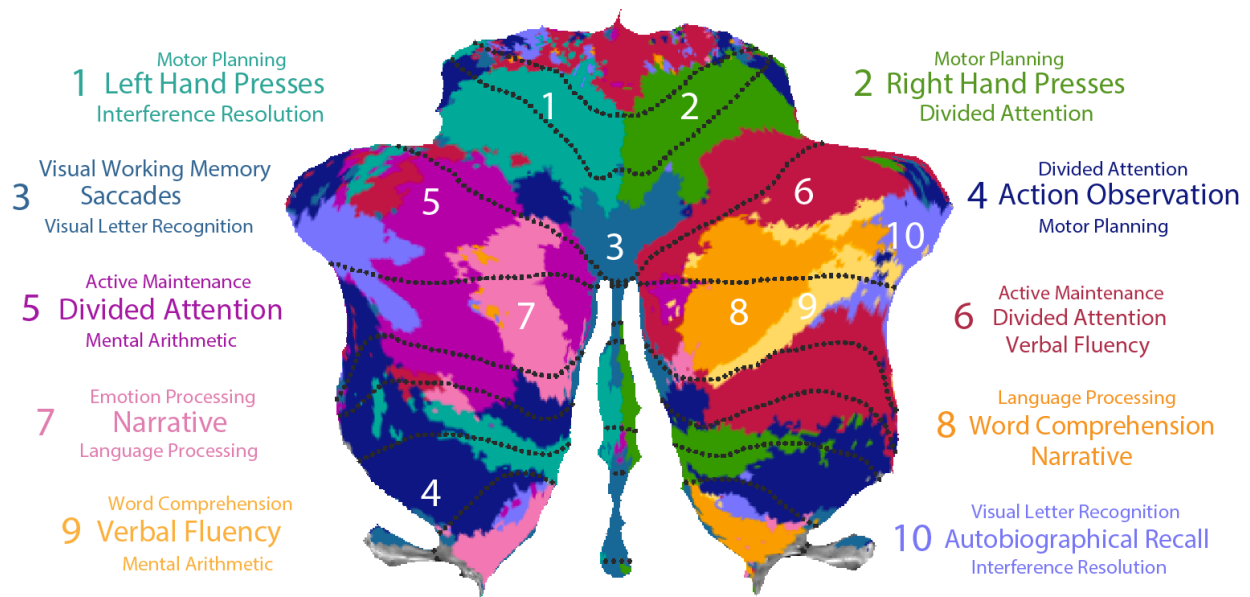


**Fig. 5: Cognitive descriptors for the ten functional regions in the MDTB parcellation.** *The three features that best characterize each region are listed. The font size indicates the strength of these feature weights.*

The dominant features describing the three motor regions (regions 1, 2 and 3) were left-hand, right-hand and saccadic eye movements, respectively. The posterior associative motor region (region 4) was driven predominantly by action observation. For the remaining regions, the dominant features related to a range of cognitive processes. Regions 5 and 6 in the mid-hemispheric aspects of Crus I/II, lateralized to the left and right hemisphere respectively, were associated with attention- and working memory-related features such as divided attention and active maintenance. More medially in both hemispheres were regions 7 and 8, best described by features associated with narrative (region 7) and word comprehension (region 8). Activity in right-hemispheric region 9, lateral to region 8, was best explained by features related to language processing (for example, verbal fluency and word comprehension). Finally, region 10, encompassing the most lateral aspects of Crus I/II was dominated by autobiographical recall. This region shows strong task-free correlations with the frontal pole and other areas related to the default mode network<sup>5</sup>. Overall, activity in the larger proportion of the cerebellum was explained by features related to cognitive, rather than motor, processes<sup>19</sup>.

## 1.5 Discussion

### 1.5.1 Summary

The aim of this study was to derive a comprehensive picture of the functional organization of the human cerebellum. To do this, a group of participants was scanned over the course of four fMRI sessions while performing a diverse MDTB. Task-evoked activation patterns were leveraged to derive a functional parcellation of the cerebellar

cortex. Using a new technique to quantitatively evaluate functional boundaries, we showed that the MDTB parcellation successfully predicted functional boundaries when tested with a novel set of tasks, outperforming existing parcellation based either on task-free fMRI data or on lobular structure.

### 1.5.2 Parcellations of the cerebellar cortex.

The lobular architecture of the cerebellum has provided, both in neurophysiological and neuroimaging studies, the primary reference for defining subregions<sup>4,20</sup>. The macroanatomical folding into ten lobules is well conserved across species<sup>4</sup> and under strong genetic control<sup>21</sup>. While the results from electrophysiological<sup>22</sup> and neuroimaging studies<sup>5,7</sup> have suggested that lobular boundaries do not demarcate functional subdivisions, we present in this study the first quantitative evaluation of this hypothesis. Indeed, lobular boundaries appear to constitute only very weak boundaries in terms of functional organization. The identified functional regions often spanned multiple lobules, with many of the boundaries traversing the cerebellar cortex along the parasagittal axis. The clear dissociation of anatomical and functional organization of the cerebellum, as revealed in the present study, questions the value of summarizing functional and anatomical data in terms of lobular regions of interest. As an alternative, we employed our task-related data to develop a parcellation that could comprehensively describe the functional organization of the cerebellar cortex. Critically, the group-based MDTB parcellation predicted functional boundaries for new tasks in the same data set, as well as for a completely separate data set (HCP task data). These findings provide a compelling demonstration of discontinuities in the functional specialization across the cerebellar cortex. Evidence from meta-analyses has suggested the existence of ‘motor’, ‘cognitive’ and ‘affective’ regions of the cerebellum<sup>8</sup>. However, it has also been suggested that functional variation across the cerebellar cortex may be best understood in terms of smooth gradients<sup>7</sup>, without definable boundaries. If this were the case, our DCBC measure, reflecting the difference of within-region to between-region correlations, would have been near zero when tested on a novel task set. Instead, the values were positive, providing a rigorous demonstration of functional boundaries in the cerebellar cortex. An open question is whether the boundaries defined through our task-based approach relate systematically to anatomical features of the cerebellum identified by molecular techniques<sup>22</sup>. Specifically, studies investigating aldolase C (zebrin II) expression in Purkinje cells in the rodent<sup>23</sup> and primate brain<sup>24</sup> have revealed a series of parasagittal zones across the cerebellar cortex. Olivocerebellar projections respect this zonal organization, with single climbing fiber inputs synapsing onto Purkinje cells that lie within a zone<sup>23</sup>. The organization of zebrin II zones remains to be established in the human cerebellar cortex. However, we suspect that the alignment with the organization observed in this study may not be very tight given that the cerebellar hemodynamic signal is primarily reflective of mossy fiber input<sup>25</sup>. Relative to climbing fibers, mossy fiber innervation patterns are likely to be patchier and more diffuse, potentially spanning multiple zonal regions<sup>26</sup>. The boundaries identified from task-free fMRI data were also able to predict task-based discontinuities. This finding is in accord with similar analyses of the cerebral cortex, demonstrating that task based activation patterns in the neocortex can be predicted to some degree by parcellations obtained from the spontaneous fluctuations in the fMRI signal during rest<sup>27</sup>. However,

our MDTB parcellation outperformed alternative task-free parcellations<sup>5,6</sup> in predicting functional boundaries for completely different tasks within the MDTB and HCP data sets. While the MDTB parcellation was based on fewer participants than the other parcellations (24 versus 1,000), our data set entailed considerably longer scanning time per participant. One notable difference between the task-free parcellations and our MDTB parcellation was that the latter assigned homologous areas in the left and right hemispheres to different regions. This was the case for regions associated with hand movements (regions 1 and 2), working memory (regions 5 and 6) and narrative comprehension (regions 7 and 8). While the parcellation suggests some degree of hemispheric asymmetry within the cerebellum, the task activity profiles between homologous regions also shared many similarities (Supplementary Fig. 7). Second, the MDTB parcellation also indicated that the areas correlated with the default mode network in task-free data could be subdivided into regions related to narrative comprehension (regions 7 and 8), language functions (regions 8 and 9) and autobiographical recall (region 10). While our group-based map could predict functional boundaries in individual participants, the finer spatial details of the functional organization were idiosyncratic for each individual. Consistent with this, the individual parcellation outperformed the group parcellation in predicting functional boundaries in that individual. Of course, individual parcellations require data collection for each participant using at least a subset of our task battery. It is worth noting that each individual parcellation was obtained on almost 3h of data per participant, an amount of scanning time that is usually not feasible or practical. In future studies, we aim to determine the required amount of data per participant, identify the best subset of tasks and explore the possibility of combining individual and group data to derive an optimal parcellation.

### 1.5.3 Insights about functional topography.

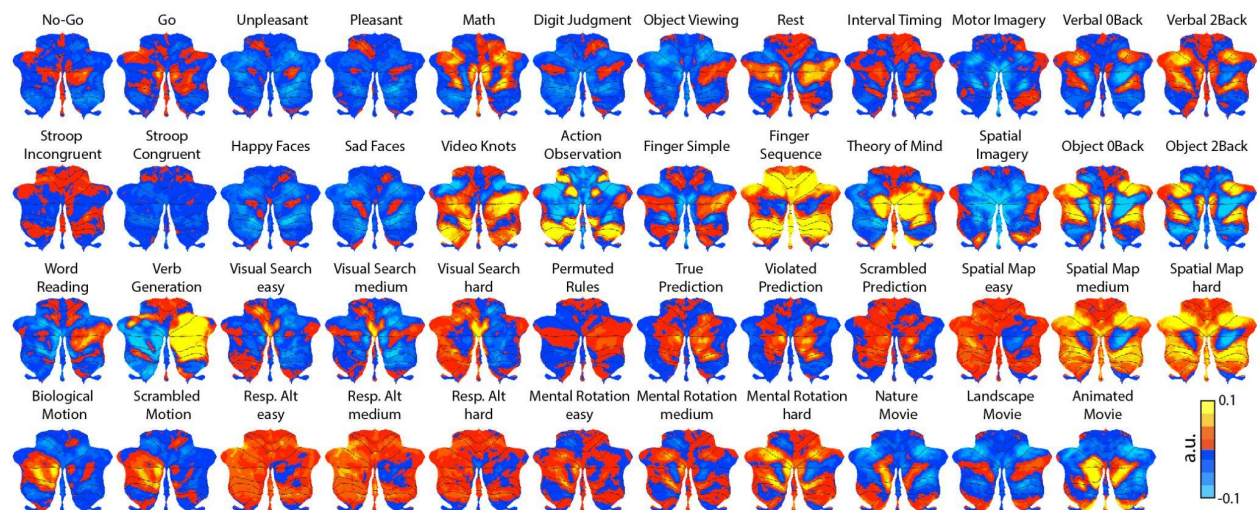
An additional advantage of a multi-domain task-based approach for mapping the cerebellum is that we can not only identify functional boundaries, but relate the activation patterns to the task requirements. For many of the tasks in our battery, the activation patterns were in accord with the results obtained in previous fMRI studies that have examined a single or limited set of task domains. Examples in the present study include working memory, hand movement, language and theory of mind tasks. By using a rich MDTB, we also identified functional regions that had not been observed or well described in previous work. A large and extended region of the cerebellar cortex was activated during action observation (region 4), surrounding the anterior, but to a much larger extent, posterior hand motor region. Interestingly, the action observation region was also activated during complex movement, as shown by the sequence production task. Taken together, these results suggest that anterior motor regions are more related to primary action execution, whereas posterior motor regions are more akin to a 'premotor' area, perhaps associated more with action planning and action comprehension. Notably, lesions limited to the posterior cerebellum rarely lead to lasting symptoms of ataxia<sup>28</sup>. A second example comes from our motor feature model, which revealed a region around vermis VI that was strongly associated with saccadic eye movements (region 3). This finding is consistent with neurophysiological data from non-human primates showing that this region is associated with oculomotor control<sup>11</sup>.

However, prior neuroimaging studies of the cerebellum have proven controversial regarding this issue. Some studies have also linked this area with eye movements<sup>12</sup>, but other studies have argued for a functional role of this region in more complex cognitive and/or affective processes<sup>7</sup>. Based on pilot work for this study and other unpublished observations, we have found it difficult to elicit any cerebellar activation with a simple saccadic eye movement task. In contrast, we observed robust activation in this area during the visual search task, even when accounting for the average number of saccades made during a 30s block. Thus, our results offer a fresh perspective on the functional role of this region, indicating that the hemodynamic signal is driven by eye movements performed under high attentional demands. The identification of this oculomotor region is also of interest given that prior studies have suggested that vermal activation in lobule VI is associated with emotional processing<sup>7</sup>. Conditions in our battery designed to engage emotional and affective processing (for example, static images of unpleasant and pleasant scenes, sad and happy faces) only weakly activated this region. Given the strong association of this region with eye movements, it may be that prior activations were more related to differences in saccadic eye movements between conditions, rather than the emotional and affective processing demands of the tasks. Further support for this hypothesis comes from task-free fMRI studies showing that the oculomotor vermis is functionally connected to visual regions of the cerebral cortex<sup>5,7</sup>. An obvious challenge for future research is to explore how variation in eye movements, affective processing and attentional demands interact in driving activation within this region.

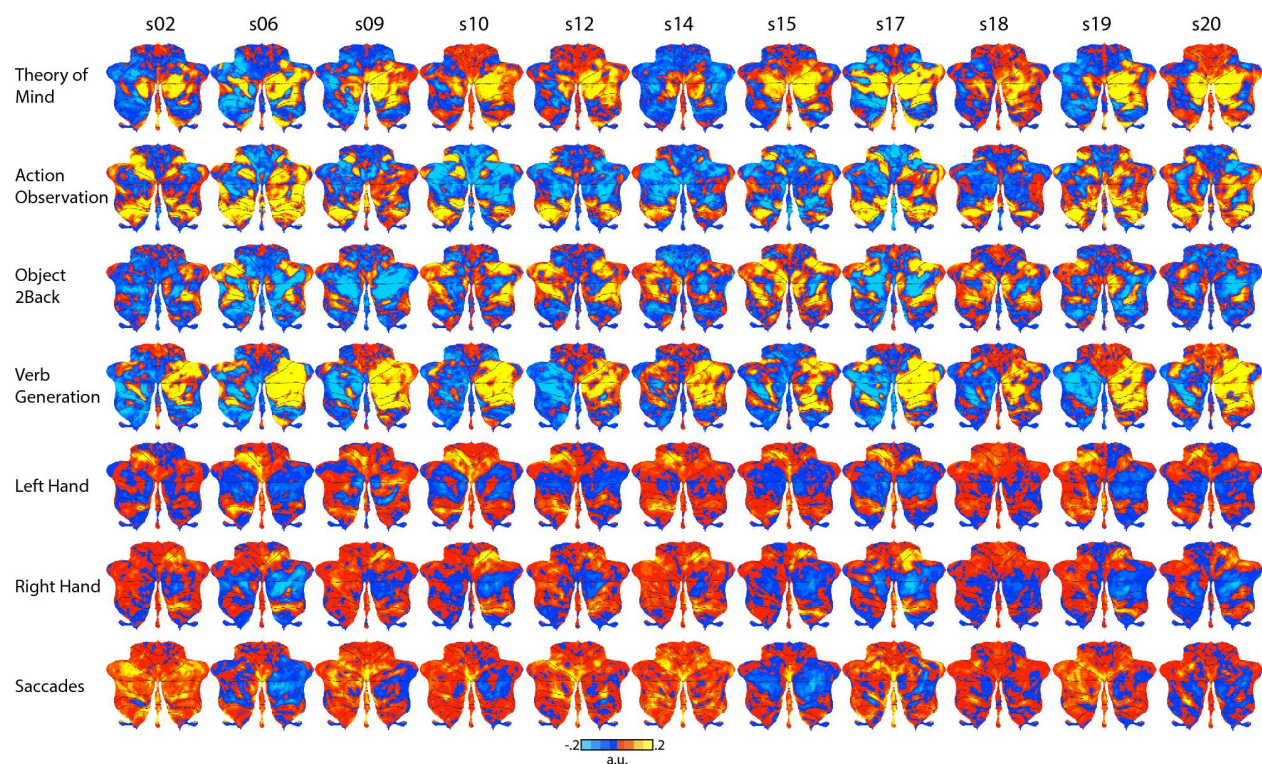
#### 1.5.4 Conclusions

This article presents a comprehensive MDTB for the human cerebellum, which is unique in its functional diversity and amount of data per individual. The group and individual task contrast maps and the group parcellations can be viewed at [diedrichsenlab.org/imaging/mdtb.htm](http://diedrichsenlab.org/imaging/mdtb.htm). We anticipate this resource will be useful for two important applications. First, the data, combined with our evaluation criterion, provide a quantitative assessment of functionally defined boundaries, something that has been absent in prior studies. Given that our acquisition parameters covered the entire brain, the methods presented in this study can be used to evaluate parcellations of the neocortex and other brain structures. Second, the novel MDTB parcellations provide an important tool to define functional regions of the cerebellum. Our analyses clearly show that the MDTB parcellations predict functional boundaries in a novel set of tasks better than existing task-free parcellations. Moreover, each region can be characterized by a rich functional task profile, allowing for a characterization of the associated cognitive processes. For future research, the parcellation and associated features can provide a useful guide in designing studies to test specific functional hypotheses and provide a reference for interpreting the results. The MDTB functional parcellation should also be of considerable utility for translational work, given the hypothesized involvement of cerebellar dysfunction in a range of neurological and psychiatric disorders<sup>29</sup>. A functionally defined parcellation can help reveal dysfunction in specific cerebellar regions and cerebro-cerebellar circuits<sup>30</sup>, providing further insight into the interaction between the cerebellum and neocortex.

## 1.6 Supplementary Material

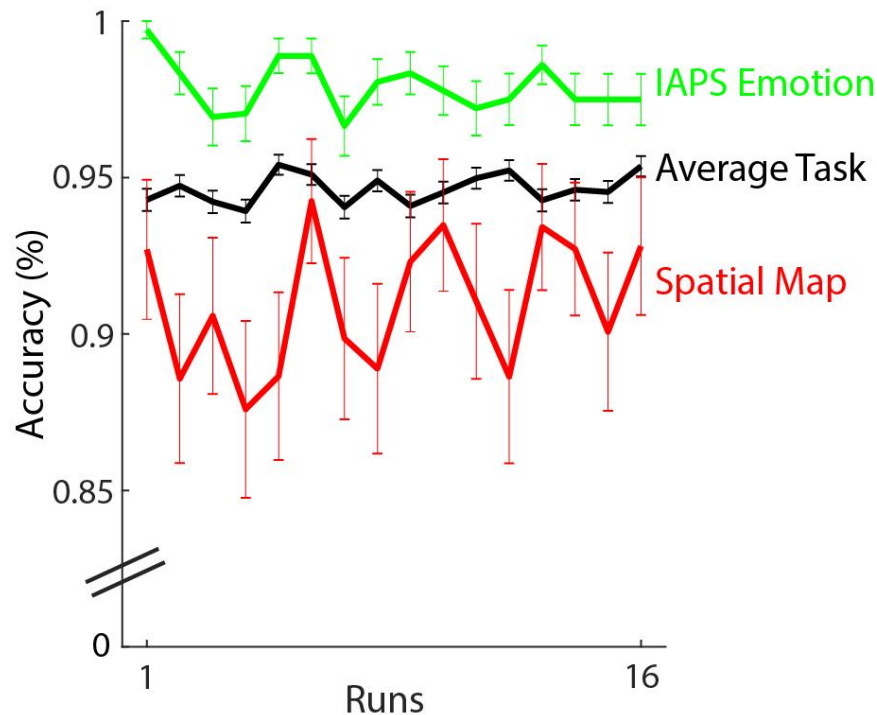


**Supplementary Figure 1.** Unthresholded, group-averaged activation maps for the 47 unique task conditions displayed on a surface-based representation of the cerebellar cortex<sup>14</sup>. All activations are calculated relative to the mean activation across all conditions. Red-to-yellow colors indicate increases in activation and blue colors indicate decreases in activation. Activity is normalized by the root-mean-square-error of the time-series fit for each voxel.

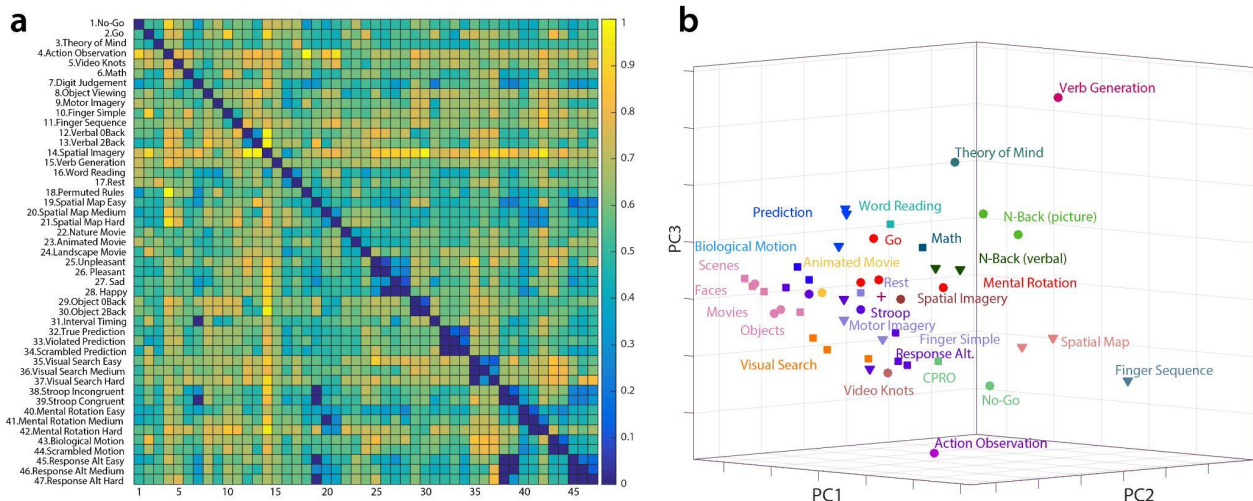


**Supplementary Figure 2.** Unthresholded, individual activation maps for 4 representative tasks and motor feature maps for 11 representative participants.

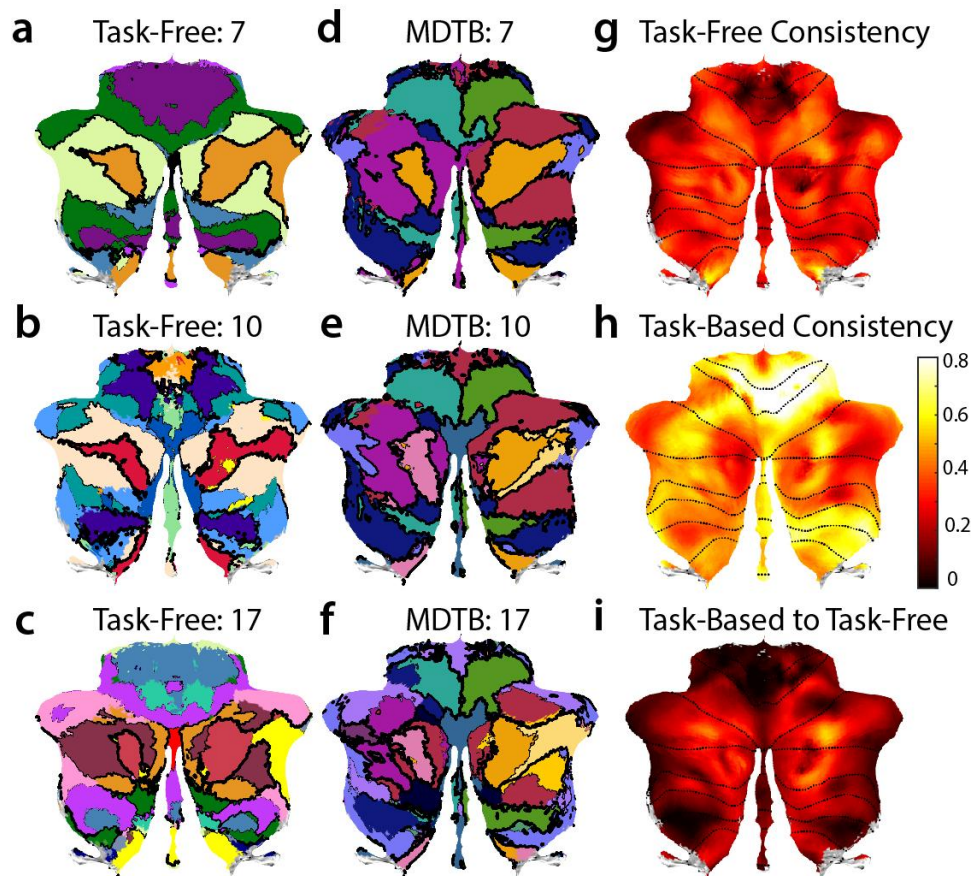
All activations are calculated relative to the mean activation across all conditions. Red-to-yellow colors indicate increases in activation and blue colors indicate decreases in activation. Activity is normalized by the root-mean-square-error of the time-series fit for each voxel.



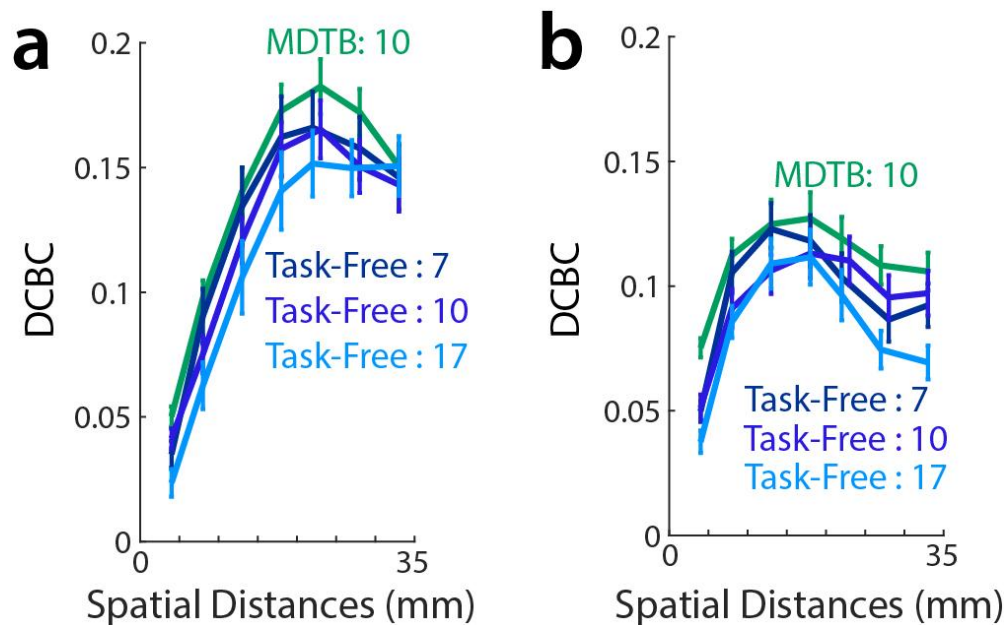
**Supplementary Figure 3. Stability of task performance.** Percent accuracy, averaged across two scanning sessions, each composed of eight runs. Average across all tasks is shown in black. Poorest performance was on the spatial map task (red line) and best performance was on the IAPS emotion task (green line). Error-bars indicate between-subject ( $N=24$ ) standard error.



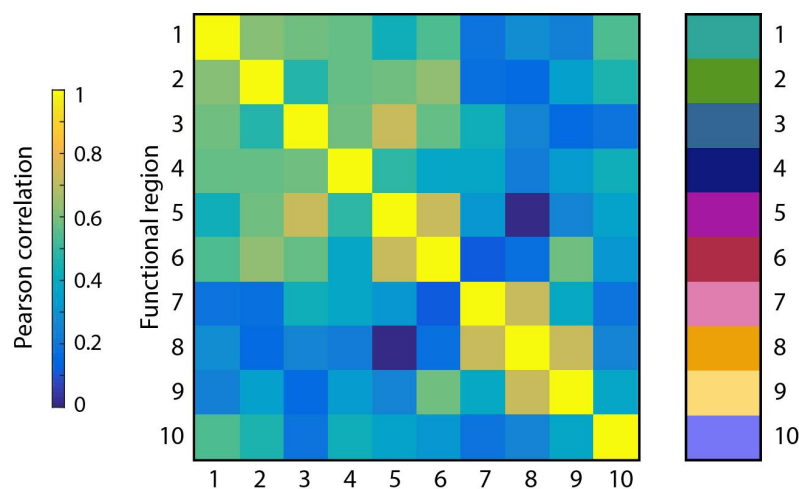
**Supplementary Figure 4. Representational task space for 47 unique task conditions.** (a) Group-averaged representational dissimilarity matrix (RDM) data for the unique 47 task conditions. Shared tasks are averaged across the four scanning sessions. Dark blue represents low dissimilarity between pairwise task-evoked activity patterns while high distances (bright yellow) represent high dissimilarity between pairwise task-evoked activity patterns. Thresholded values are shown below the diagonal (dark blue cells indicating pairwise comparisons between task conditions were not significant ( $p < .001$ , e.g., pleasant and unpleasant scenes). (b) A multi-dimensional scaling plot (MDS, using first three PCs for display purposes), showing the relative similarity of the task-evoked activity patterns after correction for activity related to basic motor output. Hierarchical clustering was applied to the tasks, with colors in both the RDM and MDS indicating cluster membership.



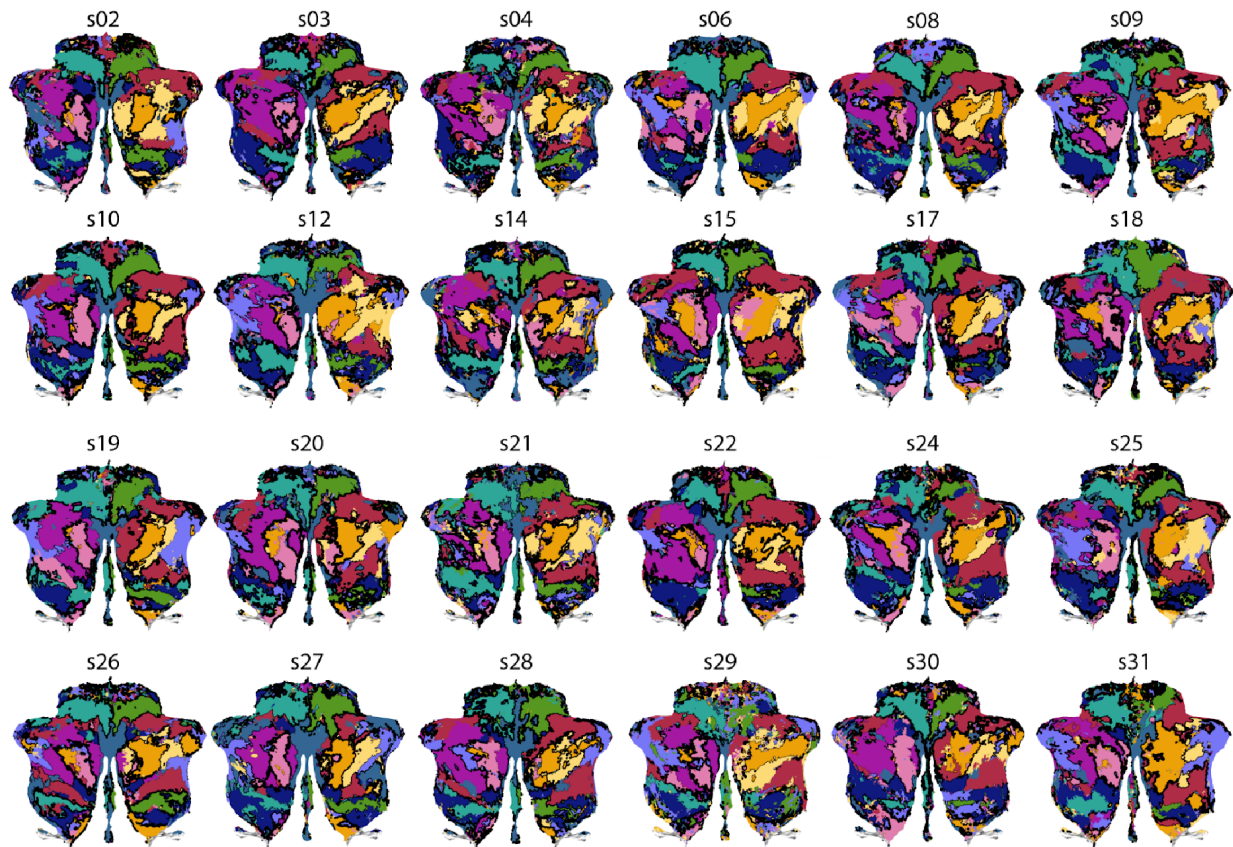
**Supplementary Figure 5. Comparison of task-based and task-free parcellations.** 7, 10, and 17 region parcellations derived from task-free HCP (a-c) and MDTB (d-f) data. (g) Average Rand coefficient between task-free parcellations, computed locally (1cm sphere) around each cerebellar voxel. (h) Average Rand coefficient between MDTB parcellations. (i) Average difference of Rand coefficients for the MDTB and task-free parcellations.



**Supplementary Figure 6. Cross-validated evaluation of MDTB parcellation on a subset of 7 tasks, selected to be most dissimilar to task conditions included in the data set.** For comparison purposes, task-free parcellations are evaluated on the same tasks. (a) MDTB parcellation trained on Set A and evaluated on 7 tasks from Set B (Mental Rotation Easy, Mental Rotation Medium, Mental Rotation Hard, Spatial Map Medium, Spatial Map Hard, Animated Movie, and Nature Movie). (b) MDTB parcellation trained on Set B and evaluated on 7 tasks from Set A (Sad Faces, Interval Timing, Go, Theory of Mind, Word Reading, Motor Imagery, Math). Error-bars indicate between-subject standard error (N=24).



**Supplementary Figure 7. Pearson correlation between the task-profiles of the 10 regions of the MDTB parcellation.** The values in the correlation matrix are scaled between 0 (blue) and 1 (yellow). The bar on the right denotes the colors of each of the 10 regions (see Fig 5).



**Supplementary Figure 8. Ten-region MDTB parcellation for 24 individual participants.** The DCBC for each boundary is visualized by the thickness of the black lines.

| Task Name                                 | Task Description                                                                                                                                | Datas<br>et | Conditions                             | Hand<br>Assign<br>ment |
|-------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-------------|----------------------------------------|------------------------|
| Object<br>Viewing                         | Passive viewing, pictures of objects and a checkerboard pattern.                                                                                | A           |                                        | None                   |
| Motor<br>Imagery <sup>40</sup>            | Imagine playing a game of tennis.                                                                                                               | A           |                                        | None                   |
| Stroop <sup>41</sup>                      | 3AFC, indicating color of stimulus word (3 colors), comparing conditions in which color-word mapping is congruent or incongruent (Stroop task). | A           | Congruent<br>Incongruent               | Both                   |
| Verbal<br>Working<br>Memory <sup>42</sup> | 2AFC, indicating if current stimulus in stream of letters matches letter displayed two items previously                                         | A           | Verbal<br>2-Back+<br>Verbal<br>2-Back- | Left                   |

|                                     |                                                                                                                                                                                                                                |       |                                            |       |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|--------------------------------------------|-------|
|                                     | (2-back). Target can be present (+) or absent (-).                                                                                                                                                                             |       |                                            |       |
| Interval Timing <sup>43</sup>       | 2AFC, indicating if a tone is short (100ms) or long (175ms)                                                                                                                                                                    | A     |                                            | Right |
| Arithmetic <sup>44</sup>            | 2AFC, indicating if simple multiplication equations (e.g. $2 \times 7 = 14$ ) are correct or incorrect. For control task, participants view a series of four numbers and indicate presence/absence of target number (e.g., 1). | A     | Math<br>Digit<br>Judgment                  | Right |
| IAPS affective <sup>45</sup>        | 2AFC, indicating if picture (scenes, animals, foods) is pleasant or unpleasant.                                                                                                                                                | A     | Pleasant<br>Scenes<br>Unpleasant<br>Scenes | Left  |
| IAPS emotion <sup>45</sup>          | 2AFC, indicating if picture depicts sad or happy face.                                                                                                                                                                         | A     | Happy Faces<br>Sad Faces                   | Right |
| Go/No-Go <sup>46</sup>              | Go-NoGo task with positive (Go) or negative (No Go) words.                                                                                                                                                                     | A     | Go<br>No-Go                                | Left  |
| Theory of Mind <sup>47</sup>        | 2AFC to indicate if short story contains true or false belief (Theory of Mind task)                                                                                                                                            | A & B |                                            | Left  |
| Rest                                | Passive viewing of fixation cross.                                                                                                                                                                                             | A & B |                                            | None  |
| Object Working Memory <sup>42</sup> | 2AFC, indicating if current stimulus in stream of objects matches object displayed two items previously (2-back). Target can be present (+) or absent (-).                                                                     | A & B | Object<br>2-Back+<br>Object<br>2-Back-     | Right |
| Verb Generation <sup>48</sup>       | Verb generation task requiring covert responses to visually-presented nouns, either repeating the stimulus (Read) or generating a verb associated with the noun (Generate).                                                    | A & B | Verb<br>Generation<br>Word<br>Reading      | None  |
| Spatial Imagery <sup>40</sup>       | Imagine walking from room to room in childhood home, with a cue specifying the path to be taken (e.g., "Imagine walking from the kitchen to the bedroom, stopping to look around at different rooms").                         | A & B |                                            | None  |

|                                                         |                                                                                                                                                                                                                                                           |       |                                                                |       |
|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|----------------------------------------------------------------|-------|
| Motor Sequence <sup>49</sup>                            | 6-element sequence, either requiring one key press with each of six fingers (bimanual) or repetition of a single key press with one finger (unimanual left or right).                                                                                     | A & B | Finger Sequence<br>Finger Simple                               | Both  |
| Action Observation <sup>50</sup>                        | Passive viewing of videos of knots being tied, learning the name of the knot (presented at top of screen) for a latter recall test.                                                                                                                       | A & B | Video Actions<br>Video Knots                                   | None  |
| Visual Search <sup>51</sup>                             | 2AFC, indicating if target stimulus ("L") is present among distractors ("T"), with varying set size (4, 8, 12).                                                                                                                                           | A & B | Small (4)<br>Medium (8)<br>Large (12)                          | Left  |
| Spatial Map                                             | Memorize a spatial mapping of numbers (either, 1, 4, or 7) for subsequent recall                                                                                                                                                                          | B     | Easy (1)<br>Medium (4)<br>Hard (7)                             | Both  |
| Mental Rotation <sup>52</sup>                           | Mentally rotate target object to determine whether it can be brought into alignment with baseline object. Difficulty is measured by angular disparity between target and baseline image. Stimuli were obtained from Ganis and Kievit (2015) <sup>53</sup> | B     | Easy (0)<br>Medium (50)<br>Hard (150)                          | Right |
| Biological Motion <sup>54</sup>                         | 2AFC to identify intact point-light walkers (either happy or sad) or scrambled walkers (fast or slow). Stimuli obtained from Troje et al. (2017) <sup>54</sup>                                                                                            | B     | Biological Motion<br>Scrambled Motion                          | Right |
| Concrete Permuted Rules Operations (CPRO) <sup>55</sup> | Apply task-rule set (logic, sensory, & motor rules) to two consecutively presented stimuli (rectangles: either red or blue, vertical or horizontal)                                                                                                       | B     |                                                                | Both  |
| Word Prediction <sup>56</sup>                           | 2AFC task to indicate if five sequentially-presented words comprise a semantically meaningful sentence. Stimuli obtained from D'Mello et al. (2017) <sup>57</sup>                                                                                         | B     | True Prediction<br>Prediction Violated<br>Prediction Scrambled | Left  |

|                                     |                                                                                                                        |   |                                    |      |
|-------------------------------------|------------------------------------------------------------------------------------------------------------------------|---|------------------------------------|------|
| Response Alternatives <sup>58</sup> | Execute a fast motor response to an imperative signal (white cross) that appears in one of 1, 2, or 4 primed positions | B | Easy (1)<br>Medium (2)<br>Hard (4) | Both |
| Nature Movie <sup>15</sup>          | Passive viewing of a nature clip of kickboxing kangaroos, taken from “Planet Earth II: Islands”                        | B |                                    | None |
| Animated Movie <sup>15</sup>        | Passive viewing of an emotional love story between two characters from the Pixar movie “Up”                            | B |                                    | None |
| Landscape Movie <sup>15</sup>       | Passive viewing of an aesthetically-pleasing clip that depicts a diverse scenery, taken from Vimeo                     | B |                                    | None |

**Table S1.** Task set description for all 26 unique tasks and 47 unique conditions. Tasks that require overt motor responses are executed either with the left, right, or both hands.

# Chapter 2: A task-general connectivity model reveals variation in convergence of cortical inputs to functional regions of the cerebellum

## 2.1 Abstract

While resting-state fMRI studies have provided a broad picture of the connectivity between human neocortex and cerebellum, the degree of convergence of cortical inputs onto cerebellar circuits remains unknown: Does each cerebellar region receive input from a single cortical area or convergent inputs from multiple cortical areas? Here we use task-based fMRI data to build a range of cortico-cerebellar connectivity models, each allowing for a different degree of convergence. We compared these models by their ability to predict cerebellar activity patterns for novel task sets. Models that allow some degree of convergence provided the best predictions, arguing for convergence of multiple cortical inputs onto single cerebellar voxels. Importantly, the degree of convergence varied across the cerebellum with the highest convergence observed in areas linked to language, working memory, and social cognition. These findings suggest important differences in the way that functional subdivisions of the cerebellum support motor and cognitive function.

## 2.2 Introduction

The last 30 years has witnessed a paradigm shift with regards to cerebellar function, with broad recognition that this subcortical structure is engaged in many aspects of human cognition. Since the first report of cerebellar activation in a semantic retrieval task<sup>1</sup>, thousands of neuroimaging papers have reported cerebellar recruitment during a broad range of tasks that cannot be attributed to the motor demands of these tasks. Functional interpretations include hypotheses concerning how the cerebellum may facilitate attentional shifts<sup>2</sup>, stimulus-response mapping<sup>3</sup>, higher order rule processing<sup>4</sup>, verbal working memory<sup>5</sup>, language<sup>6</sup>, and social cognition<sup>7</sup>. This body of work has produced functional maps of the cerebellum that depict the association of particular cognitive processes with different subregions of the cerebellum<sup>8</sup>.

Given the relative uniform cytoarchitecture of the cerebellar cortex, it is assumed that differences in function mainly arise from variation in the input to the cerebellum. Trans-synaptic tracing methods employed in non-human primates studies have revealed extensive reciprocal connections between many frontal and parietal areas and the cerebellum<sup>9,10</sup>. These studies have highlighted the closed-loop nature of these connections, with each (neo-)cortical region projecting to a specific cerebellar region, and receiving input from the same area<sup>11</sup>. In humans, resting state functional connectivity analyses have revealed a set of cerebellar networks, each one associated with a specific cortical network<sup>12–14</sup>.

An important unanswered question is whether each cerebellar region receives input from a restricted cortical region or whether it receives convergent input from

multiple cortical regions. Providing an answer to this question has important implications for our understanding of cerebellar function. An architecture marked by a one-to-one relationship between cortical and cerebellar regions would suggest that the function of each cerebellar region is to fine-tune the dynamics in its cortical input. In contrast, a convergent architecture would suggest that subregions within the cerebellum integrate information across disparate cortical regions and may coordinate their interactions. Indeed, recent work in the rodent brain has suggested convergence of mossy fibers from diverse sources onto the same cerebellar region<sup>15,16</sup>, or even onto the same granule cells<sup>17</sup>. Furthermore, the pattern of cortico-cerebellar convergence may vary across the cerebellar cortex, similar to how cortical areas show considerable variation in the degree to which they serve as points of convergence from other cortical regions<sup>18–20</sup>.

In the current study we introduce a novel approach to study the pattern of connectivity between the cerebral cortex and cerebellum. We employed an fMRI data set obtained while human participants completed a large task battery that was designed to engage cognitive processes across a broad range of functional domains (e.g., visual cognition, memory, attention, cognitive control). Using this rich data set, we built models that predicted the cerebellar activity based solely on cortical activity, with the models varying in terms of the degree of convergence of cortical inputs onto each cerebellar area. To evaluate the models, we examined how well each model predicted data obtained from different tasks and/or different participants. These analyses reveal a novel picture of cortico-cerebellar connectivity, one in which the degree of convergence was higher in cerebellar regions associated with more complex cognitive functions.

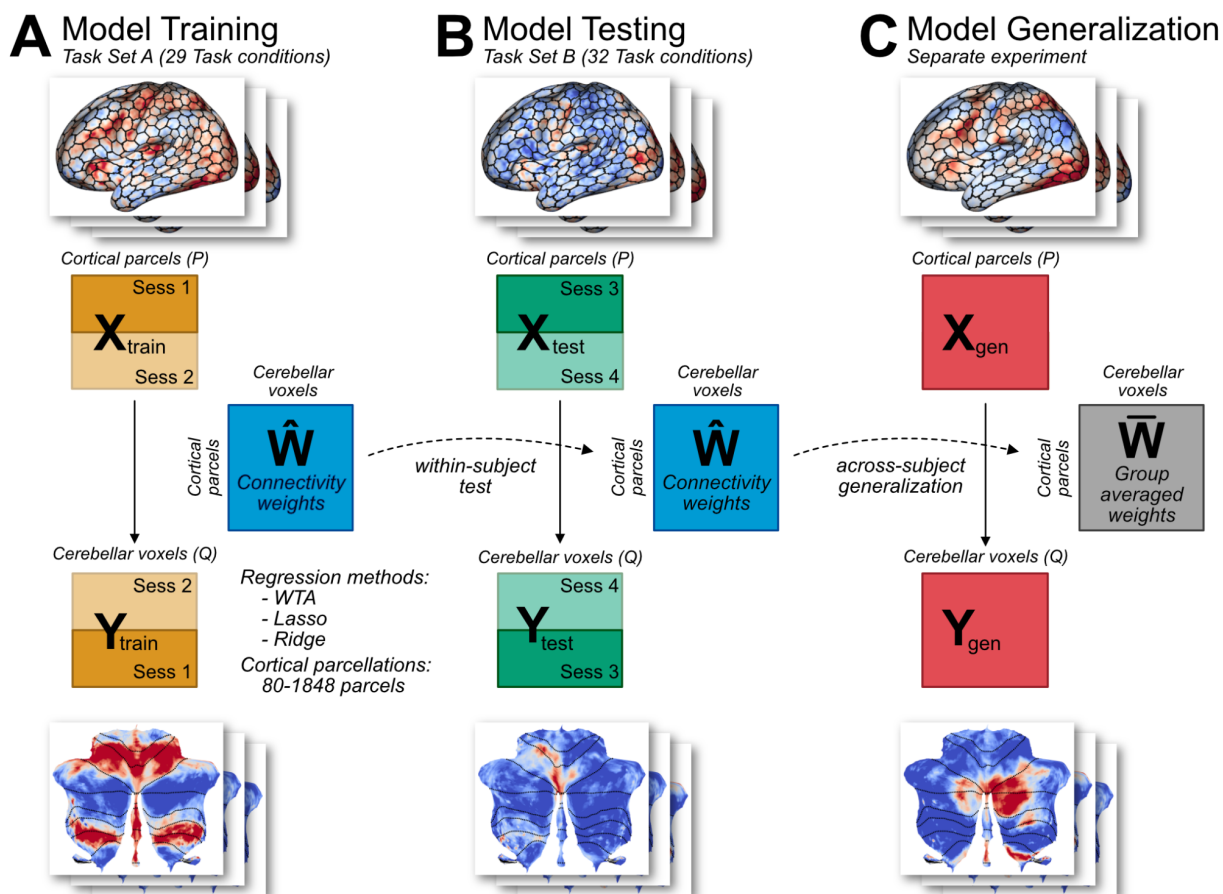


**Fig. 1 | Connectivity model training and evaluation.** (A) Models were trained on task set A of the multi-domain task battery (MDTB<sup>8</sup>). Model hyperparameters were tuned using 4-fold cross validation. Three types of models were tested (WTA, Lasso, Ridge), each with 7 cortical parcellations of different granularity. (B) Models were used to predict the cerebellar activity solely from cortical activity patterns when participants performed task set B (including both novel and common tasks). To avoid the influence of shared noise correlations across cortex and cerebellum, the models were trained and tested by using cortical and cerebellar activity patterns from different sessions (see methods). (C) As a test of generalization, the models were used to predict cerebellar activity from cortical data obtained from a separate experiment (King et al., unpublished data).

## 2.3 Results

### 2.3.1 Overview

We compared three models of cortico-cerebellar connectivity by using a region-to-region predictive modeling approach<sup>21,22</sup>. For each model, the cortical activity pattern for each task was used to predict the activity pattern in the cerebellum (Fig. 1). We chose this direction of prediction, as the cerebellar BOLD signal overwhelmingly reflects mossy-fiber input, with minimal contribution from cerebellar output neurons, the

Purkinje cells<sup>23,24</sup>. As a model of sparse connectivity, we used a Winner-Take-All (WTA) model that imposed the strong constraint that only a single cortical region could be used to predict the activity in each cerebellar voxel. The other two models allowed for some degree of convergence. Models estimated with Lasso Regression (L1 regularization) found sparse solutions, minimizing the number of cortical inputs by setting to zero those weights that make negligible contribution to the predicted activity. Models estimated with Ridge Regression (L2 regularization) allowed for a wide distribution of inputs, keeping each weight as small as possible. For the input to each model, we used a set of cortical parcellations that varied in terms of their level of granularity.

We trained the models to predict the activity in each voxel of the human cerebellum as a linear function of cortical activity (Fig. 1A), using cortical and cerebellar fMRI data obtained from 24 participants who performed a large task battery over four scanning sessions. The battery was composed of 26 tasks that could be further subdivided into 47 conditions. Set A consisted of seventeen tasks presented in an interleaved fashion in the first two sessions. Set B consisted of 17 tasks, 8 of which had been part of Set A, presented in an interleaved fashion in the second two sessions. We used Set A to estimate the connectivity weights  $\hat{W}$  for each model. These weights were then used to make predictions for the cerebellar activity patterns from Set B (Fig. 1B). Importantly, the model predictions relied on a single set of connectivity weights across all tasks, and therefore were based only on the cortical activity patterns without reference to any features of the tasks themselves. In a final generalization test, the models were used to predict cerebellar activity patterns from new participants who were tested on a set of novel tasks (Fig. 1C). Overall, the training and evaluation was designed to determine the best task-general model of cortico-cerebellar connectivity.

To validate that this approach enables us to distinguish between different forms of cortico-cerebellar connectivity, we conducted a range of model recovery simulations. Using the measured cortical activity for each participant, we generated artificial sets of cerebellar data imposing either a one-to-one or a many-to-one mapping. Following the procedure used with the real data, we trained the different models on Task Set A and tested them on Task Set B. The simulations (Extended Data Fig. 1) showed that the WTA performs best if each cerebellar voxel is connected to only one cortical parcel, whereas ridge regression performs better if there is substantial convergence, with Lasso performing in between the two methods. Thus, despite the considerable covariance between cortical parcels, the modeling approach was able to distinguish between different forms of connectivity.

### 2.3.2 Cortico-cerebellar connectivity is best captured by models with convergence

We first compared the different models, asking how well they predicted activity patterns obtained when the same participants were tested on task set B (Fig. 2A). Models allowing for some degree of convergence outperformed the WTA model. This advantage was observed across all levels of cortical granularity; indeed, the prediction performance for the Ridge, Lasso, and WTA models was relatively independent of granularity. Averaged across all levels of granularity, the Ridge models outperformed the

WTA models ( $F_{1,23} = 47.122, p < .001$ ). Post-hoc tests revealed that this advantage for the Ridge model was consistent across all levels of granularity, starting with a parcellation of 80 regions ( $t_{23} = 5.172, p < .001$ ). We also found a significant difference in predictive accuracy between the Ridge and Lasso model ( $F_{1,23} = 15.055, p < .001$ ). Post-hoc tests showed that there was no significant difference in predictive accuracy at the lowest level of granularity ( $t_{23} = 1.279, p = 0.213$ ), whereas the difference was significant for the finer parcellations (all  $t_{23} > 10.937, p < .001$ ).

The mean predictive accuracy of the Lasso and Ridge models was 0.257 (Fig. 2A). There are two issues of note here. First, the models are being used to predict activity of individual voxels, without any smoothing across cerebellar voxels. Thus, the upper bound for the predictive accuracy is limited by the reliability of the cerebellar test data. The correlation of the measured cerebellar activity patterns across the two sessions of Task Set B was, on average,  $r = 0.51$  (SD = .102). Reliability was fairly consistent across the cerebellum (Fig. 2C), with some decreases in lobules I-IV. This dropoff likely reflects the fact that our battery only included hand and eye movements, and did not activate the lower body representation that is prominent in this area. Second, predictive accuracy is also limited by the reliability of the cortical data that was used to make the prediction. This is especially a limiting factor for WTA models that use fine granularity since the predictions will be based on data obtained from a small cortical region.

Given these issues, we calculated a noise ceiling for each model, taking into account the reliability of the cerebellar data, the reliability of the cortical data, and the effect of granularity (see Methods, Extended Data Fig. 2). As an unbiased comparison of the model predictions, Fig. 2B re-plots the predictive accuracy of each model normalized by its noise ceiling. As with the original analysis, the Ridge model significantly outperformed the WTA model ( $F_{1,23} = 16.49, p < .01$ ), and in post-hoc tests, the advantage was especially pronounced for finer parcellations (1848 regions;  $t_{23} = 5.073, p < .001$ ). Overall, the noise ceiling calculation showed that the Ridge model was able to predict approximately 45% of the systematic variance of the cerebellar activity patterns across tasks (average  $R^2 = 0.67^2$ ). While the predictive accuracy (Fig. 2D) was best in anterior motor regions, it was reasonably high across the entire cerebellar surface.

The predicted and observed activity patterns for two exemplary tasks (Fig. 2E) demonstrate the quality of these predictions at the group level. Both for the spatial working memory and social cognition tasks, the connectivity model predicted the pattern of task activity with a high degree of fidelity. This is especially compelling for the spatial working memory task as the training set did not include a task with similar characteristics.

To ensure that our results were not biased by the use of an essentially arbitrary parcellation of the neocortex, we repeated the analysis using a range of published functional parcellations. For example, we used a 7-network cortical parcellation based on resting state fMRI data<sup>25</sup> to train and test the three models. Here, too, the WTA model was inferior to the Ridge models ( $t_{23} = 2.956, p < .01$ ), with no performance difference between Ridge and Lasso models,  $t_{23} = -1.235, p = 0.229$ . The same pattern

held when we used other common functional parcellations of the neocortex (Extended Data Fig. 3).

In summary, these results demonstrate that models which entail some degree of convergence from the neocortex to the cerebellum outperform a model in which cerebellar activity is based on input from a single cortical region. This conclusion holds across a broad range of cortical parcellations<sup>26</sup>.

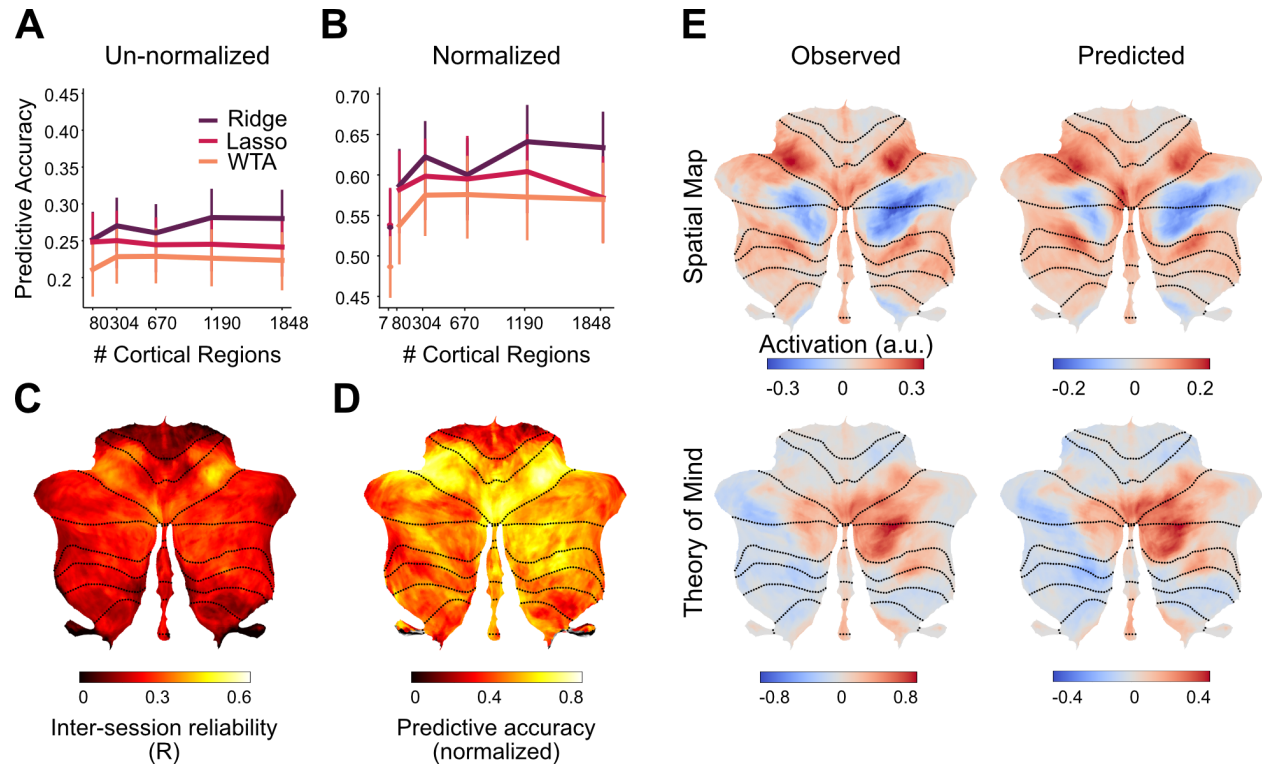


**Fig. 2 | Performance of cortico-cerebellar connectivity models.** (A) Predictive accuracy (Pearson correlation) of the Ridge, Lasso, and WTA regression models for the test data of task set B. (B) Predictive accuracy normalized to the noise ceiling based on reliability of both cerebellar and cortical data (see methods). (C) Voxelwise map of inter-session reliability of the test data. (D) Voxelwise map of predictive accuracy of the Ridge model (1848 parcels), normalized to the noise ceiling. (E) Observed and predicted activity for a novel task involving spatial working memory (Spatial Map) and a social cognition task (Theory of Mind). The spatial map task was not included in Task Set A.

### 2.3.3 Lobules VII receives more distributed input from the cerebral cortex

Given that the pattern of cortico-cerebellar convergence has important consequences for understanding the function of a cerebellar region, we were interested in knowing whether the pattern of convergence varied across the cerebellar cortex. As a first step, we tallied the number of non-zero regression weights for each cerebellar voxel as a measure of its input surface area. For this calculation, we used the Lasso model (80 cortical regions) since it uses the minimal set of cortical areas necessary for predicting

each cerebellar voxel. By setting all other weights to zero, the Lasso model allows for a quantification of the important input weights without applying an arbitrary threshold. This calculation revealed a substantial variation in the degree of convergence across the cerebellar cortex (Fig. 3a). For example, predicting the activity pattern of voxels within Crus I required inclusion of up to 10% of the cortical surface whereas predicting the activity of voxels in the anterior lobe required less than 5% of the neocortex.

To statistically analyze these data, we opted to bin the cerebellar data using a functional 10-region parcellation of the cerebellum<sup>8</sup>. This analysis confirmed that the size of the estimated cortical input area differed across functional regions ( $F_{9,207} = 7.244$ ,  $p < .001$ , Fig. 3b). The lowest level of convergence was observed for regions 1 and 2, the anterior hand regions of the cerebellum. The highest levels of convergence were found in regions of the cerebellum associated with more cognitive functions (e.g., regions 7 and 8, areas engaged during language comprehension). Noteworthy, this pattern holds for cortical parcellations of different levels of granularity (Extended Data Fig. 4), as well as for the thresholded coefficients from the Ridge regression models (Extended Data Fig. 5a, b).

As an independent method to quantify convergence, we determined the spatial spread of the inputs across the cortical surface. For this measure of dispersion, we calculated the spherical variance of the non-zero connectivity weights on an inflated model for each cerebral hemisphere (see Methods). This analysis revealed a similar pattern as seen in the surface area measure (Fig. 3c,d,  $F_{9,207} = 18.322$ ,  $p < .001$ ). For example, the cortical inputs to the hand motor regions of the cerebellum were more concentrated whereas the cortical inputs to lobule VII were more dispersed. Again, this pattern also holds for the Ridge models (Extended Data Fig. 5c, d).

The preceding analyses provide a summary of convergence from the cerebral cortex to the cerebellum. To gain insight into the spatial profile of the cortical inputs, we visualized all of the cortical weights for each of the 10 functional regions of the cerebellum (Fig. 4). As expected given the crossed connectivity between M1 and lobules IV/V<sup>9,27,28</sup>, the input to the hand regions of the cerebellum (regions 1 and 2) was centered around the contralateral primary sensorimotor cortex with some additional input from premotor and parietal cortex. Regions 3 and 4 are the other two cerebellar regions associated with motor function. Activity in region 3, the oculomotor vermis, is predicted by a bilateral set of cortical regions including the frontal eye fields (FEF), regions in the intraparietal sulcus, and extrastriate visual regions. Region 4, an area strongly activated during action observation, is predicted by a bilateral network of regions including pre-motor cortex, supplementary motor cortex (SMA) and parietal cortex. Cerebellar regions 5-10, the regions associated with more cognitive processes are predicted by a distributed set of cortical regions, with stronger input coming from the contralateral cortical hemisphere. For example, cerebellar region 5 is restricted to the left cerebellum and receives much stronger input from the right cerebral hemisphere. Overall, these findings point to significant variation in the patterns of cortico-cerebellar connectivity, with cerebellar regions in lobule VII, associated with cognitive processes such as language and working memory, exhibiting higher degrees of convergence as compared to cerebellar regions in lobule V and VIII.

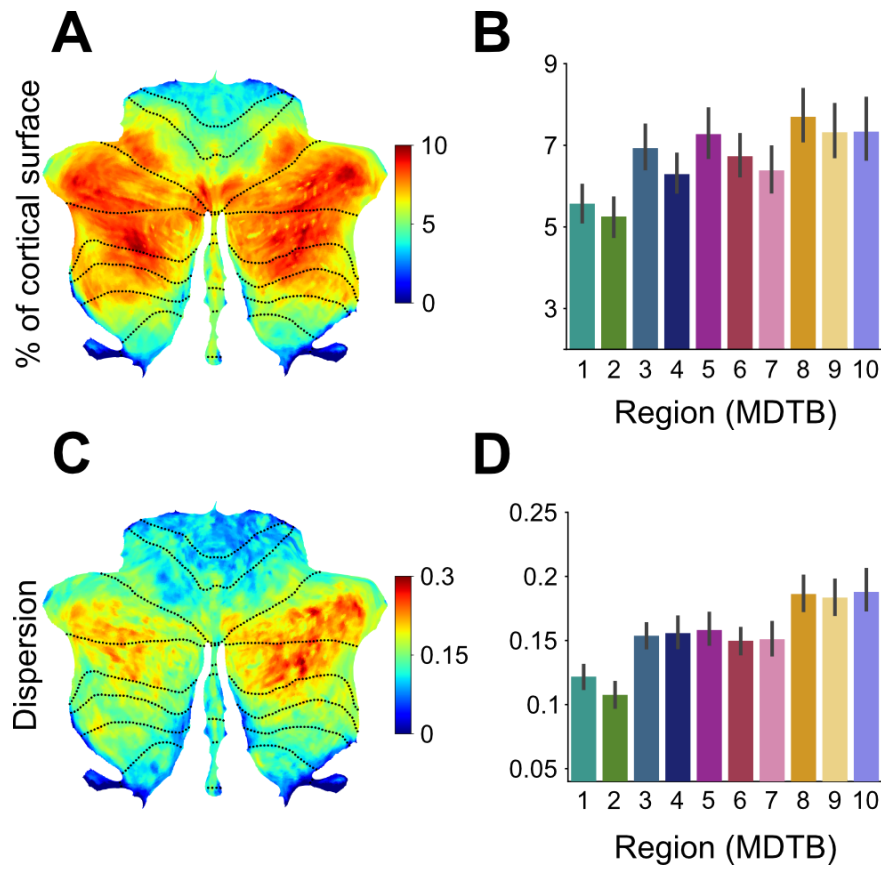


**Fig. 3 Cortico-cerebellar convergence measures using the Lasso model.** (A) Map of the cerebellum showing percentage of cortical parcels with non-zero weights for the Lasso model ( $n=80$  parcels). (B) Percentage of parcels with non-zero weights for functional subregions of the cerebellum. (C) Spherical dispersion of the connectivity weights on the cortical surface for each cerebellar voxel. (D) Average cortical dispersion for each functional subregion of the cerebellum. Error bars indicate standard error of the mean across participants. See Extended Data Fig. 5 for the same results using Ridge regression.

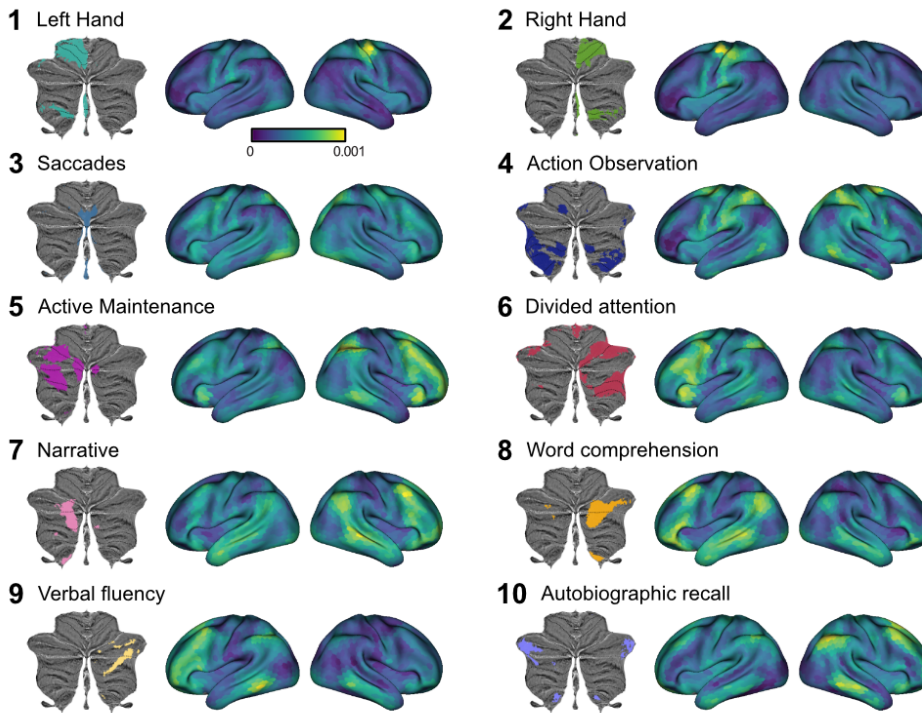


**Fig. 4 | Cortical connectivity weight maps** for the Ridge regression model with 1848 cortical parcels for each of 10 functional cerebellar regions. Each region is denoted by the most important functional term<sup>8</sup>. Results are averaged across participants. Regression weights are in arbitrary units. See Extended Data Fig. 6. for the corresponding analysis using Lasso regression.

#### 2.3.4 Cortico-cerebellar connectivity model predicts new cerebellar data

A strong test of a model is how well it predicts novel data obtained in contexts distinct from that data used to develop the model. We conducted a generalization test using data from a separate experiment involving novel tasks and new participants. The data came from a study in which participants were trained over multiple sessions on five tasks, selected to span a set of distinct cognitive domains.

All three model types (WTA, Lasso, Ridge) were evaluated, using the connectivity weights estimated from Task Set A of the main study. Because the new study had different participants, we averaged the weights for each model across the individuals in the training set. These group-averaged weights were then used to predict cerebellar activity patterns for the new data.

As shown in Figure 5, the connectivity models were generally successful in predicting cerebellar activity patterns for the new dataset. The overall pattern is quite similar to that seen in the initial model tests (in which we had used data from different tasks and sessions, but involved training and test data from the same participants). Predictive accuracy was stable across levels of cortical granularity and the Ridge model provided the best overall predictive accuracy ( $r = .657$ ) and the WTA ( $r = .352$ ) the worst

predictive accuracy. These results provide evidence that our cortico-cerebellar connectivity models capture important aspects of connectivity that are stable both across tasks and participants. Moreover, in accord with the earlier analyses based on predictions at the individual level, this generalization analysis again suggests that approximately 43% of the variation of cerebellar activity across tasks can be predicted by cortical activity alone.

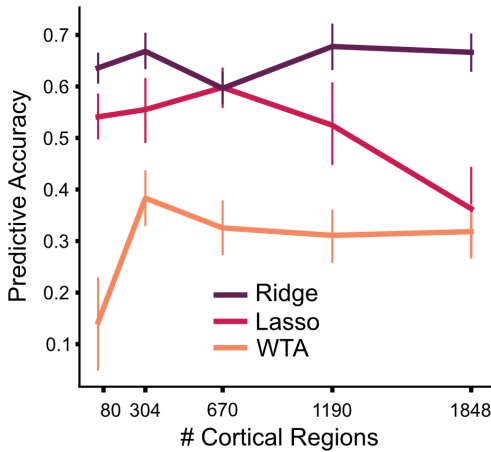


**Fig. 5 | Generalization to new dataset.** Models of cortico-cerebellar connectivity are tested on a new experiment. Each model is tested across different levels of cortical granularity. Predictive accuracy is the Pearson correlation between observed and predicted activity patterns, normalized to the noise ceiling.

## 2.4 Discussion

To date, models of connectivity between the human neocortex and cerebellum have been based on fMRI resting-state data<sup>12–14</sup>. This work demonstrates that each region of the cerebellum receives input from a distinct set of cortical regions. For example, anterior and posterior cerebellar motor regions show correlated activity with contralateral sensorimotor cortex and non-motor or “cognitive” cerebellar regions show correlated activity with specific parietal and frontal association areas.

Despite these important insights, previous work has not been designed to examine the patterns of convergence between the neocortex and cerebellum. Resting state connectivity maps are generally produced by assigning each cerebellar voxel to a single cortical network, a *de facto* winner-take-all model. In the present study, we quantified and compared models of cortico-cerebellar connectivity. The models differed in the degree of convergence of cortical inputs to each cerebellar region, ranging from an architecture constrained to a strict one-to-one mapping to architectures that allowed for distributed inputs. We evaluated these models in terms of how well they could predict cerebellar activity patterns on a novel task set. For nearly the entire cerebellum, models that allowed for some convergence predicted cerebellar activity better than the WTA model.

### 2.4.1 Convergence differs across Cerebellar Circuits

Importantly, the amount of convergence differed across the cerebellar cortex. Specifically, regions in anterior (lobules I-V) and inferior cerebellum (lobules VIII-X) were well predicted by a relatively small and concentrated area of the cerebral cortex. In contrast, regions in lobule VI and especially lobule VII required input from a larger and spatially more dispersed set of cortical regions that were primarily located in association areas of prefrontal, parietal, and cingulate cortex. This finding underscores the heterogeneity of cortico-cerebellar connectivity with some cerebellar areas functioning in nearly a 1:1 relationship with a single cortical region, whereas other areas integrate input from a more diverse set of cortical regions.

This variation bears some resemblance to a motor/cognitive gradient identified from resting state data<sup>29</sup>. However, there are a few notable exceptions. First, based on our evaluation metrics, Region 3 (oculomotor vermis) was best explained by a large and relatively dispersed set of cortical regions, including intraparietal sulcus, the frontal eye fields (FEF), and extrastriate visual areas (Fig. 4). Second, sub-regions of lobules IX, functionally associated with non-motor tasks<sup>12</sup>, are best explained by a relatively restricted set of cortical regions (Fig. 4a,c). Thus, rather than following a functional distinction, it appears that regions with strong convergence are anatomically restricted to lobules VI and VII, two areas that have disproportionately increased in size during primate evolution<sup>30</sup>.

Variation in cortico-cerebellar connectivity has important implications for theories of cerebellar function<sup>15,31,32</sup>. A cerebellar region that forms a closed-loop circuit with a single cortical region can only fine-tune the computational processes that are ongoing in its connected cortical region. In contrast, a cerebellar region that receives convergent input would be able to coordinate the interactions between dispersed cortical regions. Indeed, using a rodent model, Pisano et al. have identified cerebellar areas that project back to multiple cortical areas, providing a substrate by which the cerebellum can influence neuronal dynamics not only in a focal cortical area but also interactions within networks of cortical areas<sup>16</sup>.

### 2.4.2 Task-based vs. resting-state connectivity analyses

While our modeling approach could be applied to resting-state data, we opted to use the data from the multi-domain task battery for a variety of reasons. First, the breadth of tasks included in the battery assured that there would be reliable variation in activity across most of the neocortex and cerebellum, a necessary precondition for building a complete connectivity model. Second, the dataset allowed us to avoid biases that arise from correlated noise across abutting cortical and cerebellar regions<sup>12</sup>. We used the cortical activity patterns from one session to predict cerebellar activity patterns in a different session, relying on the fact that measurement noise is independent across sessions. Third, the MDTB data set allowed us to test the model across a broad set of tasks and mental states. While resting-state correlations are predictive of task-based activation patterns<sup>8,26,33</sup>, resting in an fMRI scanner is arguably a relatively restricted situation. Our generalization test shows that the connectivity model can robustly predict activity patterns for new tasks and participants.

### 2.4.3 Methodological limitations

Inter-region correlations of fMRI data can of course only provide indirect evidence of true functional connectivity. As such, it is important to consider methodological limitations that may influence the validity of our conclusions. From a statistical perspective, it was not clear, *a priori*, that we would have the power to distinguish between models of connectivity given that there can be substantial collinearity between different cortical regions. The model-recovery simulations (Extended Data Fig. 1a,b) suggest that the present dataset was suitable to make such inferences: Activity patterns in different cortical regions were sufficiently de-correlated, likely reflecting the use of a broad task battery. Thus, we were able to recover the correct model used to simulate the data, regardless of whether we assumed one-to-one connectivity or different degrees of convergence.

However, the simulations also indicated that the approach (and data) was not sufficient to determine the absolute degree of convergence with high confidence. For example, the size of the cortical input area for each cerebellar region differed substantially between the Ridge and Lasso regression models (Fig. 3 vs. Extended Data Fig. 5). Nonetheless, the two models result in similar predictive accuracy when using real data. Therefore, the true extent of the cortical input to a cerebellar region likely lies somewhere between the extremes provided by the Ridge and Lasso models. Importantly, this ambiguity does not impact the core observation that the degree of convergence varies systematically across the cerebellar cortex. Whether using measures based on cortical surface area or dispersion, the general picture of variation in connectivity holds for both the Ridge and Lasso models, as well as when using cortical parcellations of different granularity. Thus, we are confident that the variation in convergence reflects a stable, method-independent characteristic of the cortico-cerebellar system.

A full understanding of how the cerebral cortex and cerebellum work as a functional network will also require a detailed analysis of cerebello-cortical connections. Our decision to focus here on modeling the connectivity from cortex to cerebellum comes, in large part, from studies of cerebellar blood flow in rodents. Increases in blood flow (a major contributor to the BOLD signal) are the result of increases in mossy fiber activity, one of the primary inputs to the cerebellar cortex<sup>23</sup>. In contrast, even a dramatic increase in the firing rate of Purkinje cells does not produce a measurable change in the BOLD signal from the cerebellar cortex<sup>24,34</sup>. Thus, the cerebellar BOLD signal provides a relatively clear image of the cortical inputs to the cerebellum, but not the output of the cerebellar cortex. As such, correlations in the fMRI signal between the neocortex and cerebellum should more strongly reflect the flow of information from the cortex to the cerebellum than the reverse. A complete analysis of the entire cortico-cerebellar circuit will therefore require methods such as viral tracing techniques<sup>9,16</sup> or functional activation measures that target key nodes in the ascending pathway (deep cerebellar nuclei, thalamus).

## 2.4.4 Conclusions

Our approach was designed to identify the best task-general model of cortico-cerebellar connectivity. The connectivity model was able to take cortical activity patterns to make fairly accurate predictions of cerebellar activity, even when the data were obtained in a completely separate experiment from that used to build the model. This finding suggests that a substantial proportion of the cortico-cerebellar communication can be described by fixed information flow<sup>21</sup>; that is, each area of the cerebellum simply receives, independent of task, a copy of the neural activity occurring in the corresponding cortical areas.

However, the task-general model did not provide a perfect prediction of cerebellar activity. While these predictions may improve with more training data, we believe that there will likely be some systematic failures of the model. Such task-dependent deviations from the model prediction may offer important insights into cerebellar function, indicating that the cerebellum is more active than predicted by the cortical activity for some tasks and less for others. This pattern would suggest a “gating” of cortical input to the cerebellum, perhaps within the pontine nuclei - the main relay station of cortical inputs to the cerebellum. Rather than just transmitting cortical input to the cerebellar cortex, these nuclei may serve as an adaptive gate<sup>35</sup>, amplifying or attenuating information as a function of the relative importance of cerebellar processing for the current task.

While such task-dependent gating implies that cortico-cerebellar connectivity may be better modeled using task-1s such, this approach provides a potent new tool to test hypotheses concerning cerebellar function.

## 2.5 Methods

### 2.5.1 Multi-domain task battery

To build models of cortico-cerebellar connectivity, we used the publicly available multi-domain task battery dataset (MDTB<sup>8</sup>). The MDTB includes fMRI data from two independent task sets (A & B). Each set consists of 17 tasks, eight of which were common to both sets. The 26 unique tasks were designed to sample activity during a broad range of task domains including motor (e.g., sequence production), working memory (e.g., 2-back task), language (e.g., reading), social (e.g., theory of mind), cognitive control (no-go, stroop), and emotion (e.g., facial expression) (see Supplemental Table 1 in King et al. <sup>8</sup>).

24 participants (16 females, 8 males, mean age = 23.8) were scanned during four sessions, with task set A employed in sessions 1 and 2 and task set B in sessions 3 and 4. Within a session, there were eight 10-minute runs, with each run composed of 17 blocks of 35 s each. Each block involved a unique task and included an initial 5 s instruction period. Most tasks contained multiple conditions (i.e., different levels of difficulty), resulting in a total of 47 unique task conditions.

### 2.5.2 Image acquisition and preprocessing

All fMRI and MRI data were collected on a 3T Siemens Prisma located at the Center for Functional and Metabolic Mapping at Western University, Canada. The protocol used the following parameters: 1 s repetition time; field-of-view measuring 20.8 cm; P-to-A phase encoding direction; 48 slices; 3 mm thickness; in-plane resolution  $2.5 \times 2.5 \text{ mm}^2$ . In order to localize and normalize the functional data, a high-resolution anatomical scan (T1-weighted MPRAGE, 1 mm isotropic resolution) of the whole brain was acquired during the first session.

Functional data were realigned for head motion artifacts within each session and different head positions across sessions using a six-parameter rigid body transformation. The mean functional image was co-registered to the anatomical image and this rigid-body transformation was applied to all functional images. No smoothing or group normalization was applied. SPM12 ([www.fil.ion.ucl.ac.uk/spm/doc/spm12\\_manual.pdf](http://www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf)) and custom-written scripts in MATLAB were used to conduct data pre-processing.

### 2.5.3 General Linear Model (GLM)

To generate estimates of the activity related to each task condition, a general linear model (GLM) was fitted to the time series data of each voxel. This was done separately for each imaging run and task set (A & B). The 47 task conditions were modeled as separate regressors. Each of the 5-second task-specific instructions was modeled as a separate regressor in the GLM. For each task set, the beta weight for each task condition was averaged across the 8 runs within a session. Task set A was based on 92 activity estimates (2 sessions  $\times$  (17 Instructions + 29 condition regressors)) and task set B was based on 98 activity estimates (2 sessions  $\times$  (17 instructions + 32 conditions regressors)).

### 2.5.4 Cerebral Cortex Surface Reconstruction

For each of the 24 participants, the anatomical surfaces of the cortical hemispheres were reconstructed using the standard recon-all pipeline of the FreeSurfer package<sup>39</sup> (v. 5.0). The pipeline included brain extraction, generation of white and pial surfaces, inflation, and spherical alignment to the symmetric fsLR-32k template<sup>40</sup>. Individual surfaces were re-sampled to this standard grid, resulting in surfaces with 32,492 vertices per hemisphere.

### 2.5.5 Spatial Normalization of Cerebellar Data

The cerebellum was isolated and normalized to the high-resolution Spatially Unbiased Infratentorial Template of the cerebellum using the SUIT toolbox<sup>41</sup>. This non-linear transformation was applied to both the anatomical and functional data. Task condition activity estimates (i.e., the beta weights) were resampled to a resolution of 3 mm isotropic and resliced into SUIT space. The cerebellar isolation mask was edited to remove voxels in the superior cerebellum that abutted voxels in the primary visual cortex. Functional images were masked with the cerebellar isolation mask resulting in

activation signals that originate only from the cerebellar cortex. The cerebellar data were visualized using a surface-based representation of the cerebellar gray matter included in the SUI toolbox<sup>42</sup>. This flat map is not intended to represent a true unfolding of the cerebellar cortex, but rather provides a convenient way to visualize volume-averaged cerebellar imaging data in a 2d representation.

### 2.5.6 Connectivity Models

Model training and testing was performed on fitted time series rather than the beta weights. To obtain the fitted time series (length  $T = 9568$ ), we premultiplied the activity estimates with the first-level design matrix. This procedure reweights the activity estimates in an optimal manner, accounting for the fact that regressors for the instructions were based on 5 s of fMRI data, while regressors for the conditions were based on 10-30 s of data. This method also mitigates problems that may arise in event-related designs due to correlations of regressors<sup>43</sup>.

We normalized the time series by dividing them by the standard deviation of the residuals from the first-level GLM. The time series were then averaged for all  $Q$  cortical parcels (see next section), resulting in a  $T \times Q$  matrix ( $\mathbf{X}$ ) of cortical features. For the cerebellum, we used all  $P=6937$  cerebellar voxels (SUIT atlas space with 3 mm resolution), resulting in a  $T \times P$  matrix ( $\mathbf{Y}$ ).

Each model was trained to predict the cerebellar data  $\mathbf{Y}$  from the cortical data  $\mathbf{X}$ , using a linear model, where  $\mathbf{W}$  is the  $Q \times P$  matrix of connectivity weights, and  $\mathbf{E}$  is the matrix of measurement errors.

$$\mathbf{Y} = \mathbf{XW} + \mathbf{E}$$

When estimating connectivity models, correlated fMRI noise across cortex and cerebellum (e.g., from head movement, physiological artifacts, etc) can lead to the erroneous detection of connectivity between these two structures. This is especially problematic for the superior cerebellum and the directly abutting regions of the occipital and inferior temporal lobes. To reduce the influence of these noise correlations, we used a “crossed” approach to train the models: The cerebellar time series for the first session was predicted by the cortical time series from the second session, and vice-versa (see Fig. 1). This procedure effectively negates the influence of noise processes, given that noise processes are uncorrelated across sessions.

### 2.5.7 Cortical parcels

We created a set of models that used different levels of granularity to subdivide the cortex. Specifically, each hemisphere was divided into a set of regular hexagonal parcels, using icosahedrons with 42, 162, 362, 642, or 1002 parcels per hemisphere (see<sup>26</sup> for details). Parcels that were completely contained within the medial wall were excluded, resulting in 80, 304, 670, 1190, or 1848 parcels combined for the left and right hemisphere. Activity within a parcel was averaged across voxels for each condition, defining the  $Q$  regressors for each model.

The regular icosahedron parcellations are arbitrary in the sense that they do not align to functional boundaries in the human neocortex more than expected by chance<sup>26</sup>. We therefore also employed functionally-defined parcellations, repeating the main analyses using 12 different parcellations derived from resting state fMRI data (see Extended Data Fig. 3).

### 2.5.8 Model Estimation

We used three regression methods to estimate the connectivity matrix  $\hat{\mathbf{W}}$  at the individual participant level. In the core analyses, each of these methods was combined with the five arbitrary cortical parcellations, resulting in 15 connectivity models. Each method was selected to favor a specific pattern of cortico-cerebellar connectivity. For the **winner-take-all** models, we assumed that the time series of each cerebellar voxel is explained by one and only one cortical region. To build this model, we simply chose the cortical region with the highest correlation with the cerebellar voxel in question. The connectivity weight corresponding to the rest of the cortical regions were set to 0.

The other two methods allowed for some degree of convergence in the cortical input to each cerebellar voxel. **Lasso regression** (least absolute shrinkage and selection operator<sup>44</sup>), seeks to explain activity in each cerebellar voxel by a restricted set of cortical features. Specifically, Lasso minimizes the squared-error, plus an additional penalty calculated as the sum of the absolute values (L1-norm) of the regression coefficients:

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{XW}\|_2^2 + \lambda \|\mathbf{W}\|_1$$

In contrast, **Ridge regression** penalizes large connectivity weights, attempting to find a broad set of cortical regions to explain each cerebellar voxel. Ridge regression minimizes the following loss function:

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{XW}\|_2^2 + \lambda \|\mathbf{W}\|_2^2$$

Both Lasso and Ridge models include a hyperparameter  $\lambda$  that must be tuned in model estimation. The value of these regularization parameters was determined using gridsearch with 4-fold cross-validation on the training set. We divided the conditions of the training set into four non-overlapping subsets, reconstructed the time series using tasks from three of the four subsets and evaluated the model using the left-out subset. Extended Data Fig. 7 depicts the average predictive accuracy for the training set and the optimal value of  $\lambda$  for the Lasso and Ridge models.

### 2.5.9 Model Testing

After model training with task set A, task set B was used as the test set. We applied the same procedure to construct the  $\mathbf{X}$  and  $\mathbf{Y}$  matrices. We then used the estimated weights for each model to generate predicted voxel-wise cerebellar activity patterns from the observed cortical time series. These predictions were generated at each level of granularity for each participant. Model performance was measured by correlating

predicted and observed cerebellar time series for each voxel. As with the procedure for model training, the cortical and cerebellar time series were crossed: Cortical time series from session 3 were used to predict cerebellar time series from session 4 and vice versa. Model predictions were calculated separately for each cerebellar voxel and visualized on voxelwise maps of the cerebellar cortex.

In sum, we evaluated 15 models in which the cortical parcels were based on an arbitrary icosahedron, based on the combination of three regression methods and five levels of granularity. For the 12 functionally-defined parcellations (Extended Data Fig. 3), we limited the evaluation to the WTA and Ridge models.

### 2.5.10 Noise Ceiling

For each model evaluated in the current study, the noise ceiling quantifies the expected performance of each model, under the assumption that the estimated weights reflect the true connectivity between the cerebral cortex and cerebellum. When predicting the cerebellar activity patterns, our models are limited by two factors: Measurement noise associated with the cerebellar data and measurement noise associated with the cortical data.

To estimate these sources of noise, we used the fitted cerebellar time-series for the two sessions, assuming that each session's data are composed of the true time series  $\mathbf{Y}^*$  plus noise  $\epsilon$ :  $\mathbf{Y}_i = \mathbf{Y}^* + \epsilon$ . If  $\sigma_Y^2$  signifies the variance of the true time series, and  $\sigma_\epsilon^2$  the variance of the noise, then the correlation between the two sessions, i.e. the split-half reliability, is:

$$r(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_\epsilon^2}.$$

Similarly, the reliability of the prediction, even if we knew the true connectivity weights  $\mathbf{W}$ , is limited by the noise on the cortical data  $\mathbf{X}$ . The prediction  $\mathbf{X}_i \mathbf{W}$  has noise variance  $\sigma_{XW}^2$ , which can be estimated by calculating the split-half reliability across the two sessions:

$$r(\mathbf{X}_1 \mathbf{W}, \mathbf{X}_2 \mathbf{W}) = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_{XW}^2}.$$

Thus, for the true model the expected correlation would be:

$$r_{ceil} = r(\mathbf{X}_i \mathbf{W}, \mathbf{Y}_i) = \frac{\sigma_Y^2}{\sqrt{\sigma_Y^2 + \sigma_{XW}^2} \sqrt{\sigma_Y^2 + \sigma_\epsilon^2}} = \sqrt{r(\mathbf{Y}_1, \mathbf{Y}_2) r(\mathbf{X}_1 \mathbf{W}, \mathbf{X}_2 \mathbf{W})}$$

Because we do not know the true weights, we use the estimated weights from each model to estimate the noise ceiling. Thus, the noise ceiling is model dependent, specifying what the correlation would be if the current model was the true model.

### 2.5.11 Cortico-Cerebellar Convergence

Two measures were used to assess the amount of convergence of cortical inputs to each cerebellar voxel. For the first measure, we calculated the percentage of the cortical surface contributing to the prediction in each cerebellar voxel. We used the estimated weights from the Lasso regression model with the best unnormalized performance (80 cortical parcels). We opted to focus on the Lasso model since it sets weights that make a negligible contribution to the predictions to zero, thus identifying the sparsest cortical input that explains the cerebellar data. To determine the likely input area, we determined the number of non-zero coefficients for each cerebellar voxel, and expressed this as a percentage of the cortical surface. We also repeated these analyses using Ridge regression (Extended Data Fig. S5). Because Ridge does not result in zero coefficients, we applied a threshold, counting the number of connectivity weights with a value one standard above the mean of the weight matrix.

For the second measure, we calculated the dispersion (spherical variance) of the input weights. Using the spherical representation of the cerebral hemispheres from FreeSurfer, we defined unit vectors  $\mathbf{v}_i$ , pointing from the center of the sphere to the center of each of the cortical parcels. Each vector was then weighted by the cortico-cerebellar connectivity weight for that region  $w_i$ . All negative connectivity weights were set to zero. The spherical variance for hemisphere  $h$  across all  $Q_h$  parcels of this hemisphere was defined as:

$$var_h = 1 - \left| \frac{1}{Q_h} \sum_{Q_h} w_i \mathbf{v}_i \right|.$$

To obtain a composite measure, we averaged the variances for the two hemispheres, with each weighted by the size of the summed (non-negative) connectivity weights. For summary and statistical testing, we then averaged the size of the input area, as well as the dispersion, across cerebellar voxels within each of the 10 functionally defined MDTB regions.

### 2.5.12 Generalization to new participants

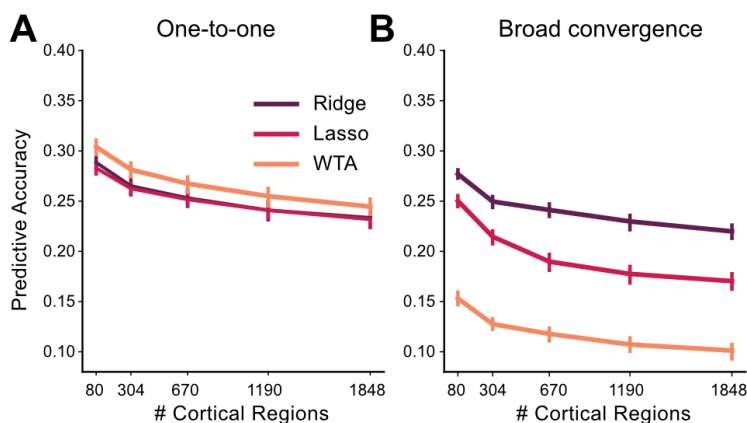
Although the models were trained and tested on distinct datasets with more than half the tasks unique to each set, the data for training and testing were obtained from the same set of participants. As a stronger test of generalization, we evaluated the models with data from a new experiment involving naive participants ( $n=20$ , 11 females, 9 males, mean age=21.3). These participants were trained to perform a 5-task battery (cognitive control, social prediction, action prediction, visual search, semantic prediction). For each task, there was an easy and hard condition (e.g., 2-back vs 0-back for cognitive control, high vs low predictability), along with the task's associated instruction. Thus, there were a total of 16 conditions when including rest (fixation). Each participant completed five sessions, with fMRI data obtained in the first, third, and fifth sessions. Within a scanning session, there were six 11-minute runs, and each run included three repetitions of each task (35 s) with a 10 s rest period prior to each 5 s instruction period.

The participants were scanned on a Siemens MAGNETOM TrioTim syngo MR B17 located in the Henry Wheeler Jr. Brain Imaging Center at the University of California, Berkeley. The protocol used the following parameters: 1 s repetition time; field-of-view measuring 20.8 cm; A-to-P phase encoding direction; 48 slices; 3 mm thickness; in-plane resolution  $2.5 \times 2.5 \text{ mm}^2$ . To localize and normalize the functional data, a high-resolution, whole-brain anatomical scan (T1-weighted MPRAGE, 1mm isotropic resolution) was acquired during the first session. fMRIPrep<sup>45,46</sup> (<https://fmriprep.org/en/stable/>) was used to preprocess the anatomical and functional data, following the same analysis procedures as conducted for the main experiment.

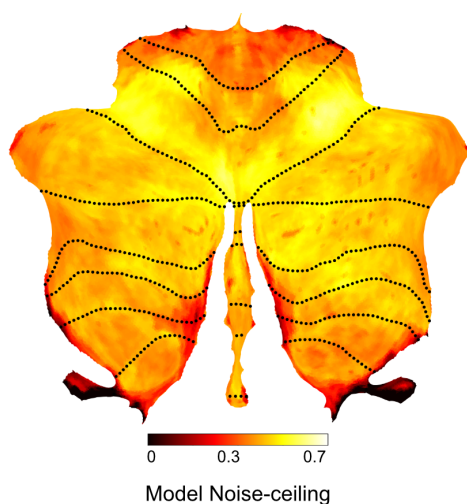
To generate estimates of activity (beta weights), a General Linear Model (GLM) was fitted for each voxel to the time series data. Each of the 16 conditions was modeled as a separate regressor, with this repeated separately for each imaging run. Nipype (<https://nipype.readthedocs.io/en/latest/>) and custom-written scripts in Python were used to estimate the beta weights from a first-level GLM.

To evaluate model generalization to this new dataset, we used the full set of cortico-cerebellar connectivity models estimated with the original data set, limiting the analysis to the regular cortical parcellations. Since the participants were different from the original data set, we averaged the model weights for the original 24 participants (i.e., 3 model types  $\times$  5 levels of granularity). Each model was then used to predict cerebellar activity from the cortical activity obtained in the new study, with these predictions compared to actual cerebellar activity at the individual level. Given that the experiment was a training study, we expected that both the true activation patterns, as well as the signal-to-noise ratio would change across sessions. We therefore determined the noise ceiling for each session separately by calculating the reliability across odd and even runs within each session. The normalized predictive accuracy was determined separately for each of the three imaging sessions, with the results across sessions averaged.

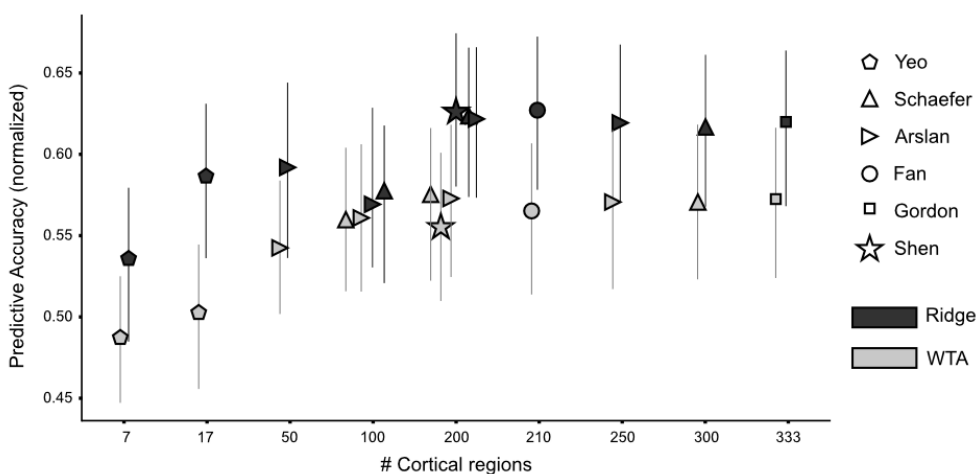
## 2.6 Supplementary Figures



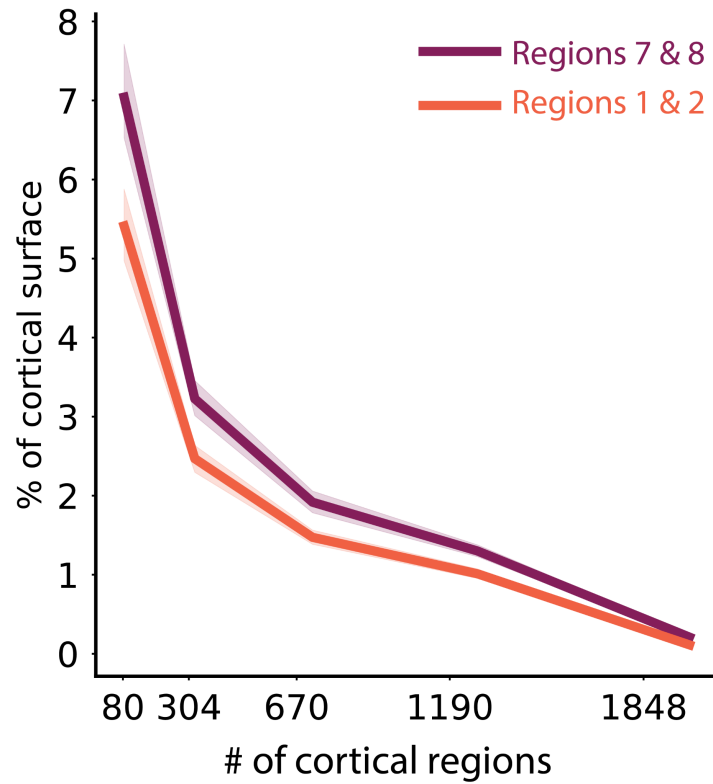
**Extended Data Fig. 1 | Model recovery simulations demonstrate the ability to identify different forms of cortico-cerebellar connectivity.** Predictive accuracy for Ridge, Lasso, and WTA models trained and tested on simulated data generated using (A) one-to-one connectivity with each cerebellar voxel connected only to one randomly selected cortical region (B) broad convergence with connectivity weights being drawn from a Gaussian distribution. Data were simulated using the observed cortical activity for task set A and B for training and test data, respectively. The signal-to-noise level for the cerebellar data was matched to the empirical data. As expected, the WTA model provided the best prediction for the one-to-one architecture and the Ridge model provided the best prediction for the convergent architecture.



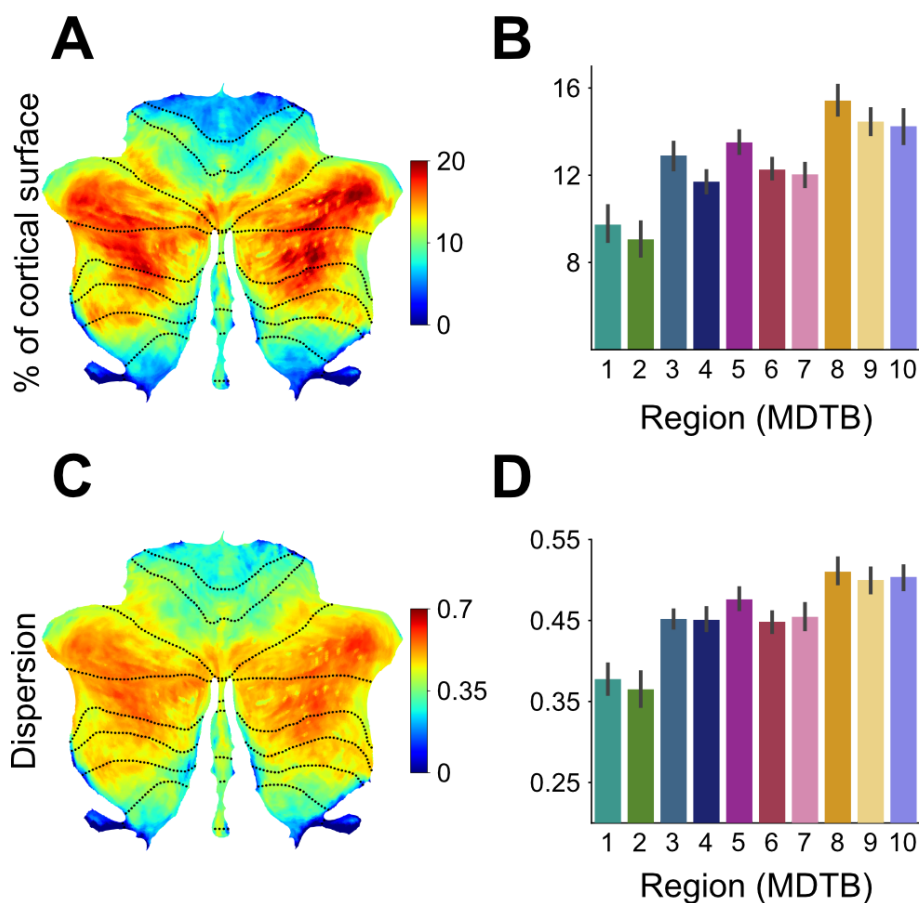
**Extended Data Fig. 2 | Noise ceiling for Ridge model.** Expected predictive accuracy assuming that the fitted Ridge model (1848 parcels) reflects the true cortico-cerebellar connectivity. The noise ceiling takes into account the reliability of the cerebellar data, as well as the reliability of the prediction based on the measured cortical data.



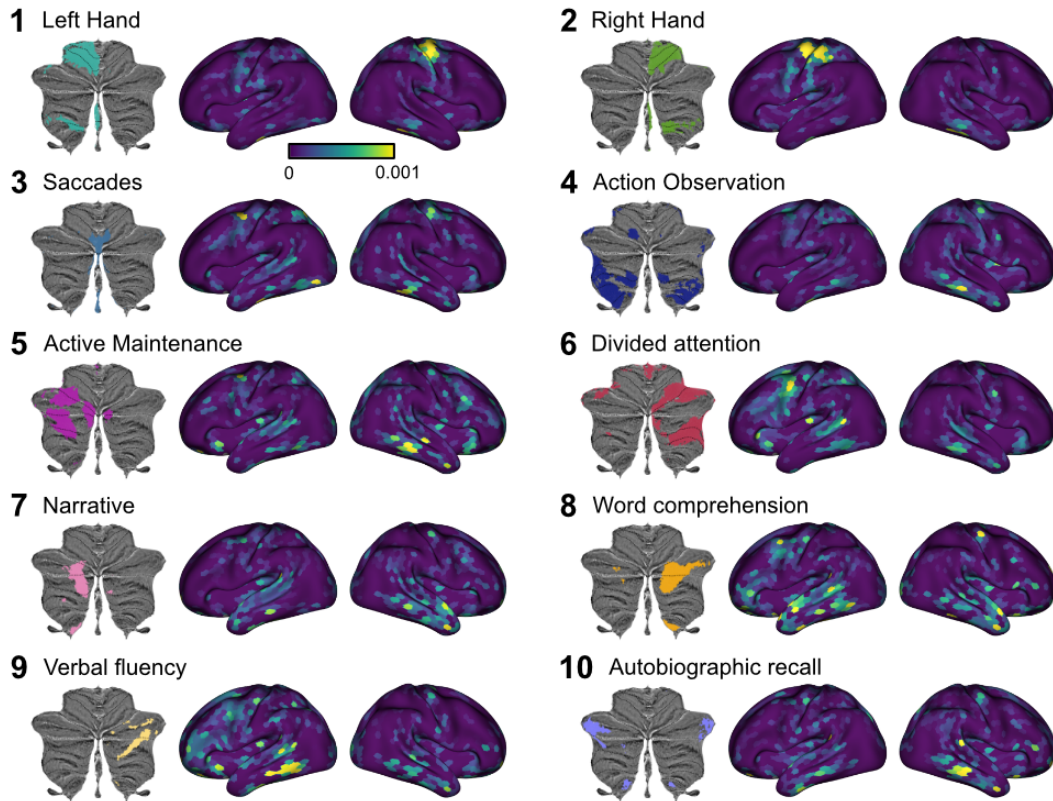
**Extended Data Fig. 3 | Predictive accuracy for Ridge and WTA models using functional cortical parcellations** (<sup>25, 47, 48, 49, 50, 51</sup>, denoted by first author). Predictive performance is normalized to the noise ceiling. The number of cortical parcels varied from 7 to 330 regions. All evaluated parcellations are available at [github.com/DiedrichsenLab/fs\\_LR\\_32](https://github.com/DiedrichsenLab/fs_LR_32)<sup>52</sup>.



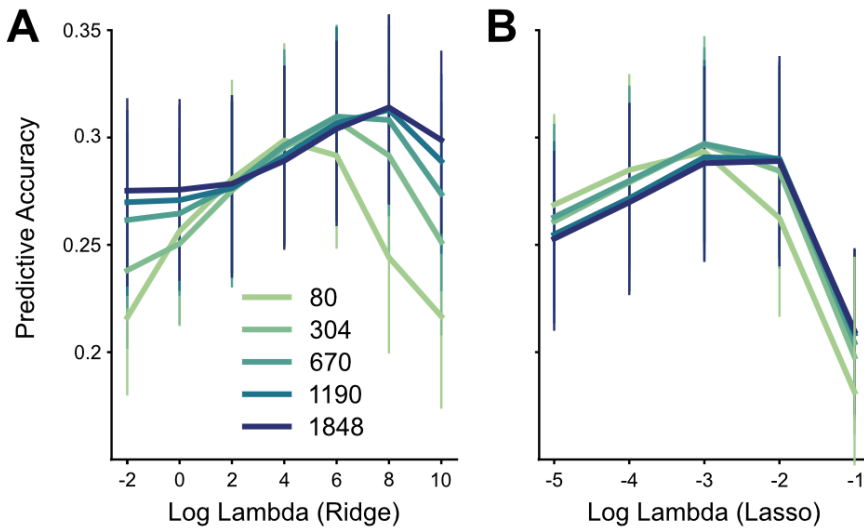
**Extended Data Fig. 4 | Percentage of cortical surface across levels of granularity for Lasso regression model.** Calculation is performed as in Figure 3b. The results are averaged across the two hand regions of the cerebellum (MDTB functional parcellation, regions 1 and 2), and regions related to narrative and word comprehension (regions 7 and 8).



**Extended Data Fig. 5 | Cortico-cerebellar convergence measures using the Ridge model.** A) Map of the cerebellum showing percentage of cortical parcels with coefficients for the Ridge model ( $n=80$  parcels) above threshold (see methods). B) Percentage of parcels with weights above threshold for functional subregions of the cerebellum. C) Spherical dispersion of the connectivity weights on the cortical surface for each cerebellar voxel. D) Average cortical dispersion for each functional subregion of the cerebellum. Error bars indicate standard error of the mean across participants.



**Extended Data Fig. 6 | Cortical connectivity weight maps for the Lasso model with 1848 cortical regions.** As in Figure 4, the results are averaged across individuals for each of the 10 functional regions defined on the MDTB data set, with each region denoted by the most important functional term. Results are averaged across participants and voxels within each cerebellar region. Regression weights are in arbitrary units.



**Extended Data Fig. 7 | Hyper-parameter tuning for connectivity models.** Predictive accuracy for Ridge (A) and Lasso (B) models using different regularization coefficients (log-lambda values) across five levels of granularity, and cross-validated over 4 folds of the training data. The WTA model did not use a hyper-parameter.

# Chapter 3: No evidence for semantic prediction deficits in patients with cerebellar degeneration

## 3.1 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 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, individuals with cerebellar degeneration showed a similar reduction in 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.

## 3.2 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, and failing to recognize the errors he was making. 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 was 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 responds to 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 and lesion (virtual or real) results. Consistent with the internal model hypothesis, increased activation is observed in 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, 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. Specifically, 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.

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.

### 3.3 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.

#### 3.3.1 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 4,000 sentences in the two databases. The 500 sentences in the Block and Baldwin database have a minimum cloze probability of 0.3. We applied this same minimum criterion to the Peelle database which excluded 1340 of the 3500 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 University 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 (Fig1 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 groups: 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.

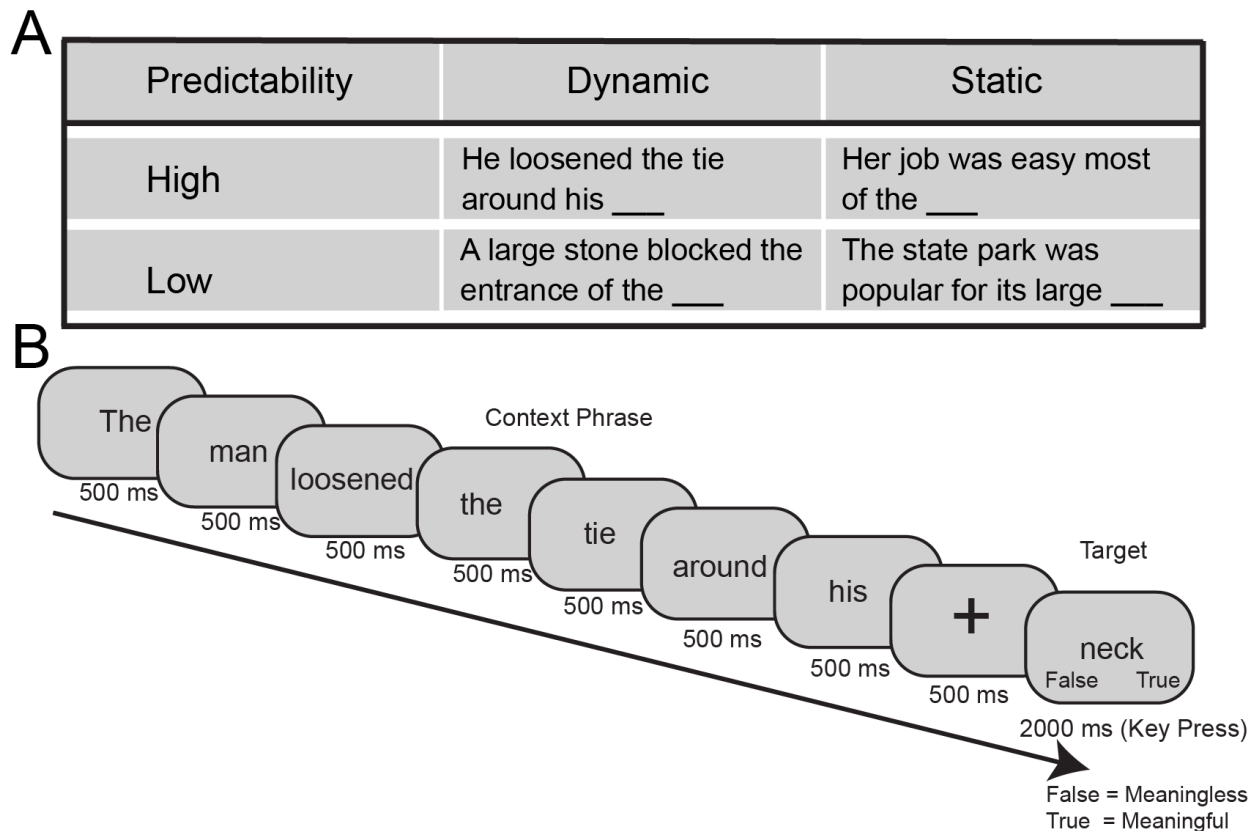


**Fig 1: Task Design.** A) Representative sentences for evaluating the internal and CoRT models. Sentence predictability is manipulated via high and low cloze probabilities, and representational transformations are manipulated via dynamic and static simulations. B) Example trial. Each word of the stem (“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.

### 3.3.2 Participants (Main Experiment)

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 the internet forum 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 in self-report) or diagnosis of 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.

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<br>(mean) | —    | 52.44 | —   | 26.39 | —     | 16.25              | —        | —        |
| control<br>(std)  | —    | 11.86 | —   | 2.72  | —     | 2.54               | —        | —        |
| patient<br>(mean) | —    | 60.5  | —   | 25.93 | 10.64 | 16.63              | —        | —        |
| patient<br>(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.

### 3.3.3 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.

### 3.4 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, and restricted 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 ( $F_{4,120} = 0.19$ ,  $p = 0.94$ ), we collapsed the data across the five blocks in the subsequent analyses.

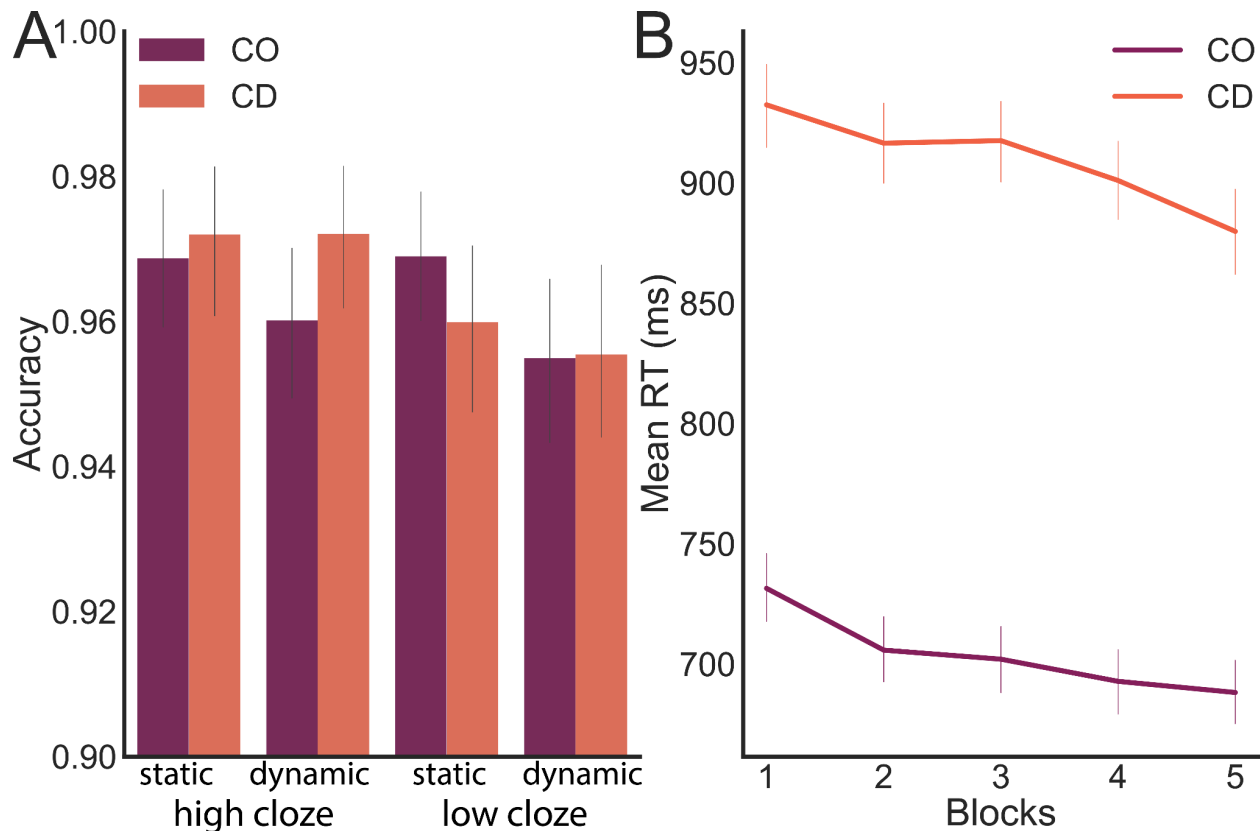


**Fig 1. Behavioral performance summary for the cerebellar degeneration (CD) and control (CO) groups.** A) Mean accuracy averaged across blocks is close to ceiling performance for both groups. B) Mean reaction time on correct trials for meaningful sentences as a function of test block (64 trials/block). 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 ( $F_{1,120} = 12.55$ ,  $p < 0.001$ ,  $\eta p^2 = 0.09$ ), 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 ( $F_{1,120} = 0.008$ ,  $p = 0.93$ ,  $\eta p^2 < .001$ ): 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 ( $F_{1,120} = 0.02$ ,  $p = 0.89$ ,  $\eta p^2 < .001$ , Fig 2C,D; Fig S1). Importantly, this factor did not interact with Group ( $F_{1,120} = 0.07$ ,  $p = 0.79$ ,  $\eta p^2 < .001$ ). 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 ( $F_{1,120} = 0.04$ ,  $p = 0.85$ ,  $\eta p^2 < .001$ )

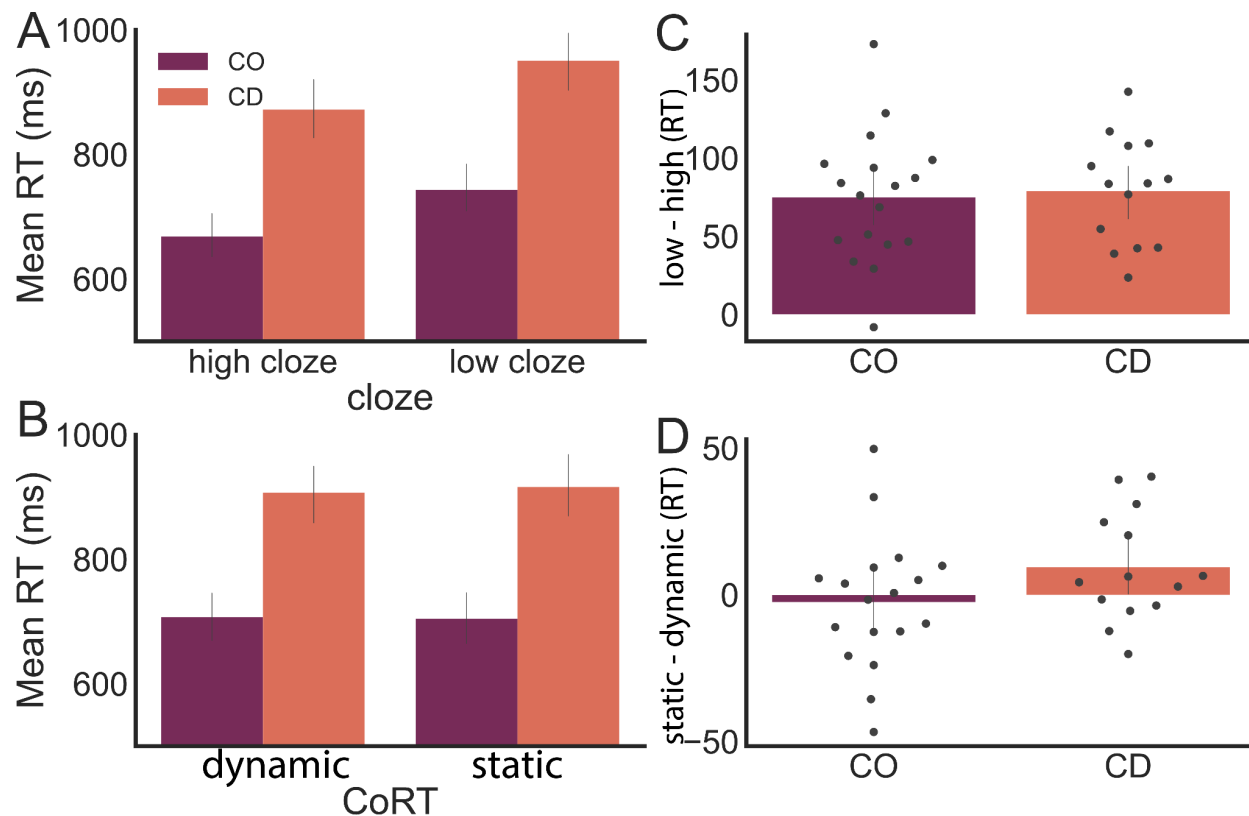


**Fig 2. No differences between the CD and CO groups in their ability to use predictive semantic information.** *A)* Mean RTs to target words as a function of cloze probability. *B)* Benefit on verification task to target words that are highly predictable (Low Cloze – High Cloze). Both the CD and CO 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,  $F_{1,120} = 12.85$ ,  $p < .01$ ). However, Trial Type did not interact with Group ( $F_{1,120} = .015$ ,  $p = .91$ ,  $\eta p^2 < .001$ ) or Prediction ( $F_{1,120} = 1.95$ ,  $p = .17$ ,  $\eta p^2 < .001$ ). Thus, 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.

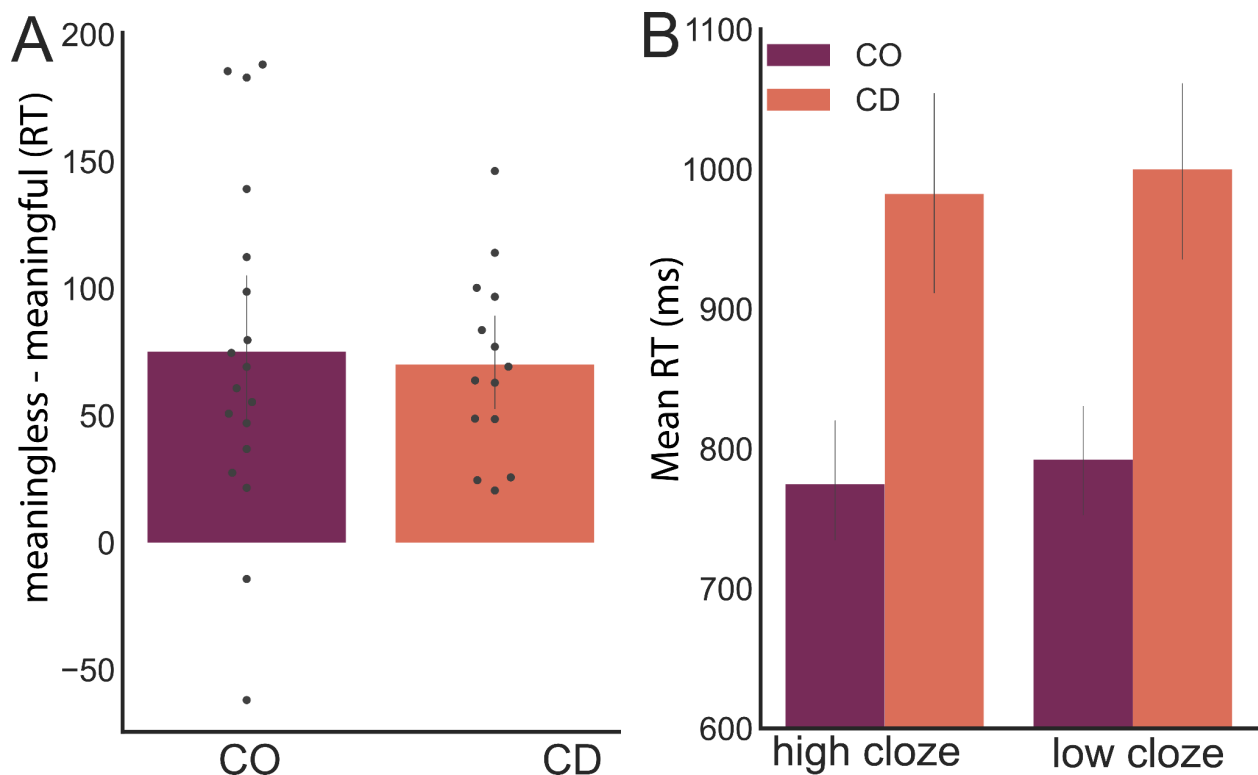


**Fig 3. No differences between the CD and CO groups in evaluating sentences that entail prediction violations.** A) RT difference between meaningless and meaningful sentences. Participants were slower when the target word did not fit with the context, but this effect was similar for the CO and CD groups. B) When the target word did not fit with the context, RT was 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.

### 3.5 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) and neurostimulation studies (D'Mello et al. 2017; Lesage et al. 2017) 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).

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 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.

### 3.5.1 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. We also recognize that prediction, at least as conceptualized here, may not be the appropriate kernel for understanding how the cerebellum supports language.

### 3.6 Supplementary Figures

| Sentence                                            | CoRT Rating | Cloze | Trial Type  |
|-----------------------------------------------------|-------------|-------|-------------|
| 1. The farmer collected wood to build a new barn    | Dynamic     | Low   | Meaningful  |
| 2. They were startled by the sudden noise           | Dynamic     | High  | Meaningful  |
| 3. Drake didn't like meat because of the taste      | Static      | Low   | Meaningful  |
| 4. His boss refused to give him a raise             | Static      | High  | Meaningful  |
| 5. The driver drove too quickly around the corner   | Dynamic     | Low   | Meaningful  |
| 6. She roasted the marshmallow over the fire        | Dynamic     | High  | Meaningful  |
| 7. His worst fear was that the spectators would boo | Static      | Low   | Meaningful  |
| 8. The patient's broken arm took a while to heal    | Static      | High  | Meaningful  |
| 9. The teacher wrote the problem on the game        | Dynamic     | High  | Meaningless |
| 10. She wanted a great job so she could get bear    | Static      | Low   | Meaningless |

**Table S1. Example sentences indicating CoRT and Cloze ratings.** Sentences 1-8 are examples of meaningful sentences, while sentences 9 and 10 are meaningless. All sentences used in the experiment (including CoRT and cloze ratings) can be downloaded [here](#).

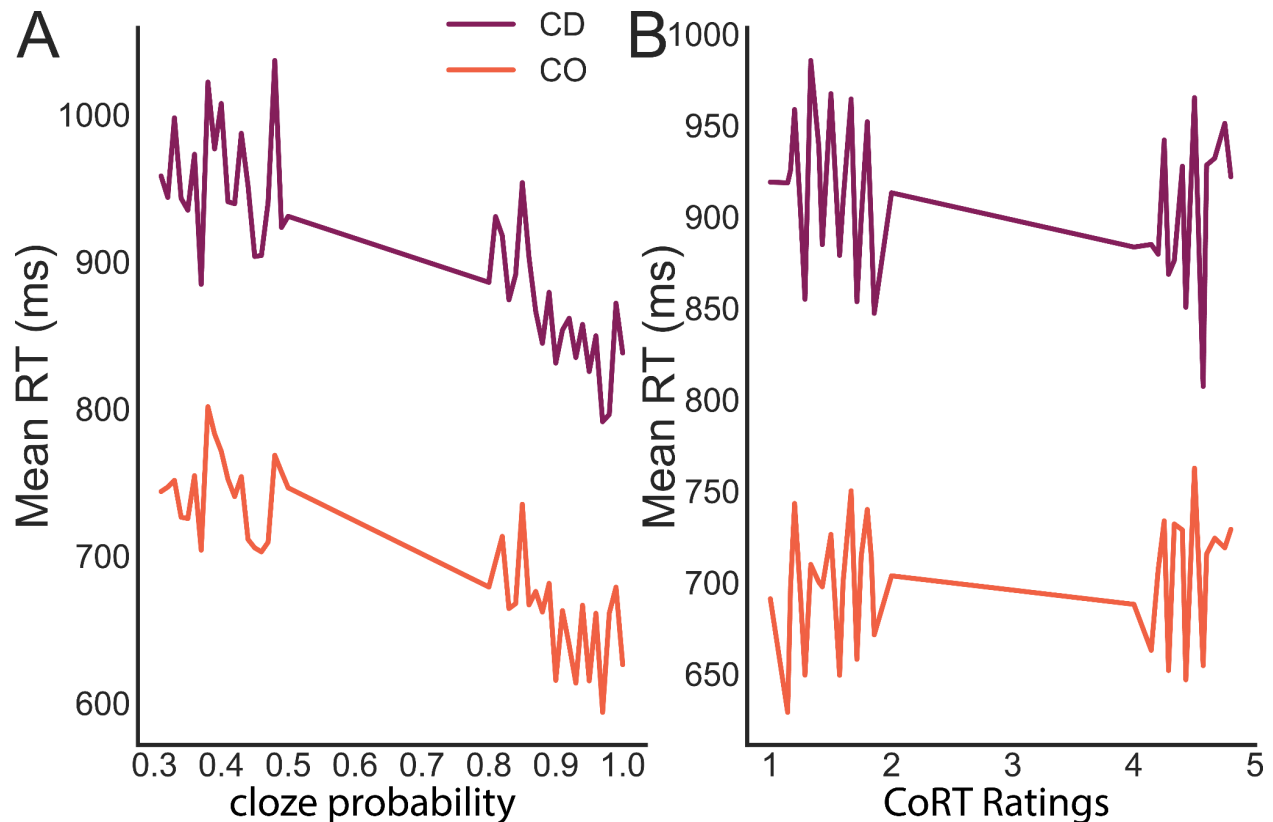


**Fig S1. Reaction time performance across cloze (A) and CoRT (B) ratings for the CD and CO groups.** Same data as shown in Figure 2, but binned as a function of (A) cloze probability or (B) CoRT rating (where high values correspond to sentences rated as most dynamic). The test stimuli were limited to sentences with a mean cloze probability of 0.3 - 0.5 (low cloze) or 0.8 - 1.0 (high cloze). Inclusion in the test set was limited to sentences with a CoRT ratings 1.0 – 2.0 (static, non-CoRT) or 4.0 – 5.0 (dynamic, CoRT)

# Chapter 4: Unique cerebellar contribution to early learning: evidence from action prediction tasks

## 4.1 Abstract

A number of theories have been proposed about how and what the cerebellum contributes to cognitive function. However, traditional neuroimaging approaches have often failed to consider how task-evoked cerebellar activation may be explained by commensurate activation in corresponding regions of the cerebral cortex. Here, we use a detailed understanding of connectivity between the neocortex and the cerebellum to test whether the cerebellum plays a mechanistic role in learning across functional domains. Specifically, we leveraged a pre-trained cortico-cerebellar connectivity model to predict cerebellar activity patterns across learning for five unique tasks: Semantic Prediction, Visual Search, Action Prediction, N Back, and Social Prediction. To determine whether there were task-specific deviations from the linear connectivity model, we compared observed and predicted cerebellar activity patterns. We found tasks that engaged dynamic action predictions were underpredicted by the connectivity model, while perceptual and working memory tasks were overpredicted, suggesting a unique functional role for the cerebellum in action prediction. Importantly, while cerebellar BOLD signal decreased across learning for all domains, the only tasks that were underpredicted by the cerebellum at each stage of learning were processes that engaged dynamic action prediction. These findings suggest that the cerebellum may play a task-specific role in coordinating cortico-cortico interactions during early stages of learning.

## 4.2 Introduction

The cerebellum receives input from almost the entire cerebral cortex (Schmahmann & Pandya, 1997), which makes it the largest bandwidth of information processing in the human brain. Inputs to the cerebellar cortex arrive mostly via the pontine nuclei through mossy fiber projections (Kelly & Strick, 2003). The classical view of the cerebellum was motor-centric, proposing that while the cerebellum receives information from many neocortical regions, it funnels back information to primary motor cortex exclusively (Allen et al. 1974). We now know, courtesy of trans-synaptic tracing studies, that there are closed anatomical loops that reciprocally connect regions in the neocortex with distinct cerebellar sub-regions (Dum & Strick, 2003; Middleton & Strick, 2001). Recently, this anatomical work has been substantiated by resting state functional connectivity (rsFC) studies that have shown how cerebellar networks are functionally connected with specific cortical resting-state networks (O'Reilly et al. 2010; Buckner et al. 2011; Guell et al. 2018).

Over the past three decades, much of what we understand about the cerebellum's contribution to cognition has come from functional magnetic resonance imaging (fMRI) studies. The advent of this research began with a study by Petersen et al. (1989), which reported that the cerebellum was activated in response to single word

processing, a distinctly non-motor task. Since this landmark study was published, hundreds of neuroimaging papers have been published demonstrating cerebellar activation in response to numerous cognitive processes including attention (Allen et al. 1997), response reassignment (Bischoff-Grethe et al. 2002), higher order rule processing (Balsters & Ramnani, 2011), cognitive control (D'Mello et al. 2020), verbal working memory (Marvel & Desmond, 2010), language (Moberget et al. 2014), and social cognition (Van Overwalle et al. 2014). While much of the progress has been piecemeal, two different approaches: meta-analysis (Stoodley & Schmahmann, 2009) and broad based task batteries (Stoodley & Schmahmann, 2012; King et al. 2019) have demonstrated a very comprehensive understanding of cerebellar functional organization.

Despite the large number of neuroimaging studies conducted to date, the role of the cerebellum in cognitive processing remains elusive. The most widely accepted theories of cerebellar operation originate from sensorimotor learning and control (Wolpert et al. 1998; Shadmehr et al. 2010). Many of these theories, most notably the internal model hypothesis (Ito, 2008), have been used to explain cerebellar fMRI activity patterns across many cognitive domains, including language (Moberget et al. 2014), and social cognition (Van Overwalle et al. 2019). In the context of motor control, the internal model has been used to explain how the sensory consequences of action are predicted by the cerebellum, and these predictions are then used to overcome time delays associated with feedback control (Wolpert et al. 1998). This idea has been extrapolated to language by suggesting that the cerebellum builds semantic models to generate expectations that are based on semantic context, for instance, to predict the next word in a sentence (Moberget et al., 2014; Lesage et al., 2017). The timing hypothesis is another popular theory that proposes a cerebellar involvement in representing the precise timing between events and motor commands (Ivry & Keele, 1989). This theory has been used to explain motor and perceptual deficits in patients with cerebellar degeneration (Ackermann et al., 1997; Spencer et al., 2003; Breska & Ivry, 2018). Theories that focus exclusively on understanding unique cerebellar contributions to cognition tend to be largely descriptive and difficult to operationalize. For example, it has been suggested that the cerebellum works as a locus of “mental coordination” (Andreasen et al., 1998), facilitating a smooth transition between sequences of events like words in a sentence (De Smet et al. 2013) or a series of social interactions (Van Overwalle et al., 2019). Similarly, Schmahmann (2004) has proposed a “dysmetria of thought” hypothesis. This idea is analogous to the process by which the cerebellum computes sensory prediction errors to make feedback corrections in sensorimotor control. Instead, this thought experiment suggests that the cerebellum produces “cognitive” or “social” prediction errors to correct errant cognitive or social processes (Stoodley & Tsai, 2021) like a “slip of the tongue” or a “social faux pas”. These metaphors are formulated on the observation that the cerebellum may facilitate, rather than generate, smooth coordination of action (Sokolov & Ivry, 2017), and this mechanistic process is likely to extend beyond sensorimotor control into cognitive function.

Another perspective is that the cerebellum calibrates action by facilitating cortical processing, rather than computing unique functions such as model-based prediction or precise timing of temporal relationships (Sokolov & Ivry, 2017). This “mechanistic” hypothesis of cerebellar operation has been studied within the context of learning, namely by measuring learning-related BOLD activity in the cerebellum while participants acquire and obtain new words (Lesage et al. 2016), automate first-order rules (Balsters & Ramnani, 2011), and learn new motor sequences (Imanizu et al. 2000). This perspective provides a control theory framework of cerebellar function (Desmond et al. 1997), suggesting that control of information processing shifts between the cerebral cortex and cerebellum during the process of learning (Ramnani, 2006). There are two ideas about shift directionality in information processing. The first proposes that the cerebellum plays an important role in “late learning”. The idea is that the cerebellum acquires learned representations from the cerebral cortex early in the learning process and then later mimics these operations with the goal of executing them in an automatic manner (Marr, 1969; Thach, 1998). The second idea supports an “early learning” hypothesis, namely that during early learning, the cerebellum works to establish patterns of communication between cerebral regions via the pontine nuclei which together are thought to entail a complex and non-linear integration of the input from the cerebral cortex (Schwarz & Thier, 1999). Analogous to the connectionist model of hippocampal-cortical systems proposed by McClelland et al. (1995), the idea is that once the mapping has been established, slower Hebbian-based learning in the cerebral cortex may consolidate cortico-cortico networks, making the involvement of the cerebellum less necessary later in learning.

Despite the appeal that many of these theories hold, fMRI research has not been very successful in identifying what the cerebellum contributes to cognitive function, beyond what would be expected from activation in corresponding cortical regions. One possibility is that the theories are difficult to operationally define and thereby difficult to test, or it may be that none of the theories developed for sensorimotor control generalize well to highly parameterized cognitive processes. However, there is another possibility, which is that classical approaches to fMRI data analysis have provided us with a skewed interpretation of how and what the cerebellum contributes to cognition. Namely, traditional fMRI approaches either ignore task-evoked activation in corresponding regions of the cerebral cortex, or at most analyze the cortical activity patterns in concert with the cerebellum. These approaches are problematic for two reasons: First, we know that the majority of the cerebral cortex projects, via the pontine nuclei, to the cerebellum (Schmahmann & Pandya, 1997), and second, we know it is highly likely that the BOLD signal measured in the cerebellum dominantly reflects mossy fiber inputs (Mathiesen et al. 2000), which carry projections from the cerebral cortex. Therefore, while traditional approaches can conclude how a cerebellar region is activated in conjunction with a corresponding cortical region, it cannot infer a unique cerebellar function relative to contributions from the cerebral cortex.

To identify unique cerebellar contributions to cognition, above and beyond what would be expected from cerebral cortex involvement, we adopted a novel approach. Specifically, we evaluated task-specific changes in cerebellar activity patterns as a function of predictions made by models trained on cortical activity patterns alone.

Central to this approach is a valid model of cortico-cerebellar connectivity that can reliably identify networks between corresponding regions of the cortex and the cerebellum. Several studies have mapped cortico-cerebellar networks using resting state data, however, by virtue of the winner-take-all method used, the networks are constrained to a “one-to-one” mapping (Buckner et al. 2011; Guell et al. 2018). In a recent study, we identified whether cortico-cerebellar connectivity operates with a less restricted mapping (King et al. 2022). We used a rich fMRI task battery (MDTB; King et al. 2019) and trained and tested different linear models to predict cerebellar activity patterns, using only cortical activation. Models that allowed for convergence performed better than more restricted models (e.g., winner-take-all model), in predicting cerebellar activity patterns.

Here, we used a pre-trained model of cortico-cerebellar connectivity to test for violations of the linear model across a number of unique cognitive tasks (Fig. 1A). Each of the tasks represents a unique functional domain established by our group in earlier work (King et al. 2019). We chose tasks with the following two criteria in mind: they could be learned across sessions, and they are known to activate both the cerebral cortex and the cerebellum. Previous work reported that a task-invariant connectivity model did not provide a perfect prediction of cerebellar activity, therefore, we expect that a model of cortico-cerebellar connectivity will fail in characteristic ways. Namely, we hypothesize that the systematic failures of this model will be task-dependent, meaning that the cerebellum will be more active than predicted by cortical activity for some tasks but not for others (King et al. 2022). We also hypothesize that there will be characteristic changes in how well the model predicts cerebellar activity patterns across learning. The “early hypothesis” proposes that the cerebellum may help to establish patterns of communication between cerebral regions early in the learning process, dropping out later in learning (Fig. 1B), while the “late hypothesis” suggests that representations learned in the cerebral cortex are acquired by the cerebellum where they are automated once learning has been established. We believe that we can subject these models to a strong test: for example if the “early” model is true, cerebellar activity early in learning should be much larger than what would be predicted from a simple linear connectivity model because the non-linear gating process of the pontine nuclei (Schwarz & Thier, 1999) would not be captured by the model; thus, learning-dependent changes resulting from gating should not be predicted by the model (i.e., cerebellar data should be underestimated early in learning). Of course, there is the possibility of a null model, which would show that while activity levels in the cerebral cortex and cerebellum vary as a function of practice, a connectivity model derived at one level of practice will prove sufficient to predict cerebellar activation at a different level of practice (Fig. 1B). This would mean that, across different stages of learning, knowing cortical activity patterns is always sufficient to predict cerebellar activity patterns.

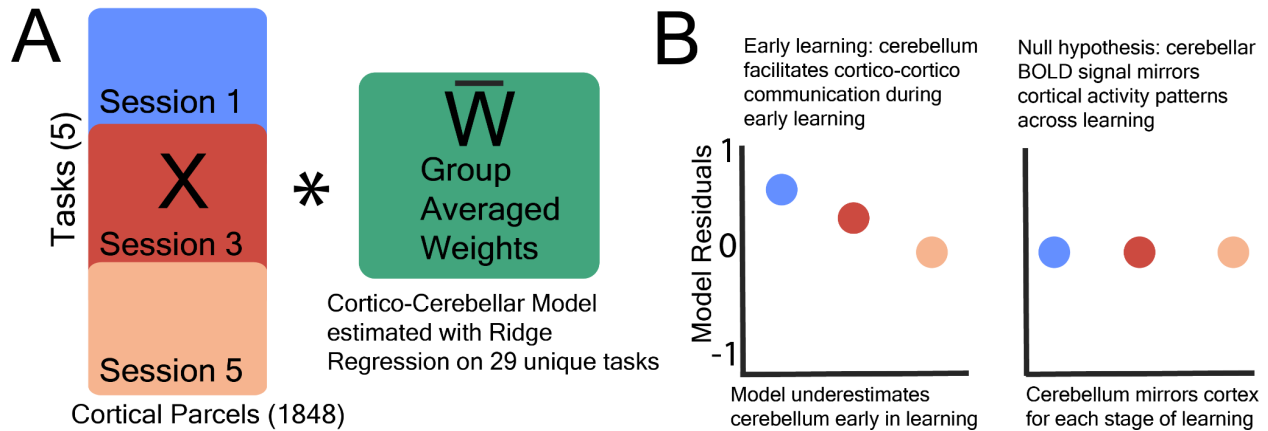


**Figure 1. Schematic of modeling approach.** A) Group weights from a pre-trained model of cortico-cerebellar connectivity (King et al. 2022) were used to predict cerebellar activity patterns of 5 unique tasks across three sessions of learning. The model was estimated with Ridge Regression to allow for convergence between the cortex and the cerebellum. B) Two hypotheses of cerebellar function (detectable with fMRI). Early learning hypothesis proposes that a cortico-cerebellar linear model will underestimate cerebellar activity during early learning, as it will not capture the nonlinear integration of input arriving via the pontine nuclei from cortex to cerebellum. Null hypothesis posits that the cerebellar BOLD signal mirrors cortical activity patterns, therefore a model trained on cortical activation patterns will always predict cerebellar activation.

## 4.3 Methods

### Participants

All participants gave informed consent under an experimental protocol approved by the institutional review board at the University of California, Berkeley. A total of 25 participants were scanned across all behavioral and fMRI sessions (5 total sessions; Fig. 2A). Three participants were excluded from the analyses because they failed to complete all 18 scanning runs. Therefore, the final sample for the analysis consisted of 22 healthy individuals (12 women, 10 men; mean age = 21.3 years, s.d. = 2.6) with no self-reported history of neurological or psychiatric illness. All participants self-reported as right-handed (Edinburgh Handedness Inventory > 40). The sample size was chosen to allow for an accurate assessment of inter-participant variability of cerebellar functional organization (King et al. 2019).

#### 4.3.1 Experimental Tasks

The dataset contained three task sets, collected at 5 different timepoints, each containing the same set of 5 tasks. The task battery was comprised of 16 task conditions, including visual search (high conjunction load, low conjunction load), working memory (N-back+, N-Back-), semantic prediction (high cloze, low cloze), social prediction (high context, low context), action prediction (high context, low context), and rest (fixation). The tasks were chosen with the following requirements: 1) they could be

learned over multiple sessions, 2) they each represented a unique cognitive domain, 3) they could be modulated for levels of difficulty, and 4) they could elicit widespread activation across the cerebral cortex and the cerebellum. We chose the tasks based on a previous multi-domain task battery in which each of the tasks (or representative tasks) demonstrated strong and distributed cerebellar activation (King et al. 2019). Three of the tasks in the current experiment were included in the original multi-domain task battery: visual search, N Back, and semantic prediction. The experimental design for each of these tasks was largely identical to the original battery. There were some minor exceptions: we used a different set of images (fractals instead of household objects) in the N Back task, and we used a different set of sentences in the Semantic Prediction task (Block & Baldwin, 2010; Peele et al. 2020), instead of stimuli by D'Mello et al. (2017). In addition, we did not include the Scrambled Prediction condition in the current version of the Semantic Prediction experiment. There were two domains in the multi-domain task battery that we wanted to include in the current task set: action observation and social cognition. However, the tasks used to represent these domains in the multi-domain task battery were deemed ineligible for the purposes of the learning taskset for two reasons: they were either naturalistic (i.e., movie-viewing tasks), or did not have a sufficient number of trials (theory of mind task). Both of these factors made the tasks unlearnable. Therefore, we designed two novel tasks to fulfill the requirements of the learning taskset. A full description of the tasks, along with the accompanying references, can be found in Supplementary Table 1. Moreover, example trials can be found for each of the tasks in Supplementary Figures 1-5.

#### 4.3.2 Construction and Evaluation of Stimuli: Action Prediction Task

To create novel stimuli for the Action Prediction task, we recorded varsity soccer players as they kicked a soccer ball at 16 unique targets (see Supplementary Figure 5). We recruited both male and female soccer players from the National Collegiate Athletic Association's (NCAA) Division I teams at the University of California, Berkeley. We hired three players (2 male, 1 female) to participate in the shoot, and they were compensated for their time. Over the course of a full day (8am - 4pm), we recorded each of the three players taking shots at 16 different targets. The targets were constructed using 8 standard issue soccer training poles (8 ft in height), and a visible number was placed at the top and bottom of each pole to construct 16 unique targets. Four of the targets were laid out in a straight line (3 ft apart) to resemble the height and width of a regulation soccer goal post size (8 ft in height x 24 ft in width), and the other four targets were placed on the outer bounds, 2 on each side of the "goal post", spaced 3 ft apart. The video camera was positioned in the middle of the "goal post", to resemble where a goal keeper would be standing before a penalty kick. The ball was placed 36 ft from the video camera (distance of a penalty kick). The video camera captured the soccer players' movements as they ran up to the starting position, and kicked the ball to one of the 16 targets. Each of the players were instructed to hit each of the 16 targets 21 times. There was variability across the players in terms of number of shots taken and their accuracy in hitting the instructed target: player 1 took 351 shots and had an average hit rate of 92%, player 2 took 340 shots and had an average hit rate of 95%, and player 3 took 320 shots and had a hit rate of 91%. We discarded shots that missed the target, however, we included shots that ended up hitting a neighboring target

(25/1011 shots), resulting in a total of 960 unique video clips. To edit and truncate the full-length clips into 1-2 sec clips, we used an open-source, free computer graphics software toolset, called Blender (Blender Foundation, 2019). We utilized Blender's built-in video sequence editor to cut our full-length stimulus videos into 1-2 sec clips. The raw videos were individually uploaded to the Blender sequencer as .mp4 files, and any residual audio was stripped from the videos. We manually toggled through clips frame-by-frame to identify the "interaction" start frame and recorded these frame numbers in a separate excel spreadsheet. The "interaction" start frame is the earliest frame where it is evident that the soccer player's foot has just touched the ball (i.e., before the ball leaves the ground). Additionally clip start and end frames were identified and manually verified under Blender's Start and End timestamp locator. The start of the clip was marked when the soccer player first started their run-up to the ball. The end of the clip was marked when the ball was no longer in the frame, and the player had completed their kick. These frames of the video were then rendered as animations and saved out as extracted 1-2 sec clips. Shots that were targeted within the "goal post" (targets 5 thru 12) were labeled as "goals" and shots that were targeted outside the "goal post" (targets 1-4 and 13-16) were categorized as "misses".

#### 4.3.3 Construction and Evaluation of Stimuli: Social Prediction Task

We created novel stimuli for the Social Prediction task by recording 2 actor pairs as they performed two different types of social interactions (hugs and handshakes) in a home setting (see Supplementary Figure 4). Both actor pairs consisted of one man and one woman. Each pairing included one of 2 of the lab's current research affiliates (female) joined by one of 2 other independent acting partners (male). Video footage was standardized as much as possible by ensuring that both actors were fully in frame (recorded above the knee), in good lighting, and had both hands visible. Videos were shot in landscape to capture the full initiation, lead up, and final interaction locations and time points. Floor filming spots were marked out so that actors consistently started the walk from the same points on opposite sides of the room and initiated their interactions in the middle of the shot (roughly 6 feet from the left and right actor starting points). The video camera was placed on a tripod perpendicular to the final interaction spot. A wide array of interaction conditions were tested by shooting various combinations of the following factors: agents involved (1 of 2 possible male-female actor pairings), label (hug or handshake), interaction initiator (female or male initiating the interaction), angle (initiator starting from the left or right of the screen), and interaction type (over or under for hug initiator).

To edit and clip the full-length videos, we used the Blender graphics application. We utilized Blender's built-in video sequence editor to cut our full-length stimulus videos into 1 second clips (Blender Foundation, 2019). The raw videos were individually uploaded to the Blender sequencer as .mp4 files, and any residual audio was stripped from the videos. We manually toggled through clips frame-by-frame to identify the "initiation" and "interaction" start frames and recorded these frame numbers in a separate excel spreadsheet. The "initiation" start frame is the earliest frame where it is evident that one or both actors are starting the interaction (ie. subtle physical signs of hand movement or arm reaching, participants are still relatively distanced on screen). The "interaction" start frame is the frame where the ending interaction actually begins

(i.e. when actors' hands or arms first physically touch in a handshake or hug; interaction hasn't yet reached its end stage). Additionally, clip start and end frames were identified and manually verified under Blender's Start and End timestamp locator. The start of the clip was marked when one of the actors first appeared on screen. The end of the clip was marked by the resolved end interaction (ie. hands made full contact for handshake, or bodies made full contact in hug). Although exact frame numbers were not uniformly quantified across all videos, they were selected based on the same video editing examples and criteria explicitly outlined in video clipping instructional documentation. These frames of the video were then rendered as animations and saved out as extracted clips.

After the extracted clips were saved post-Blender processing, they were further modified before being used in our task. Based on the time frame depicted in each clip, stimuli were categorized as one of two prediction conditions: easy or hard. The actual final interactions of the original extracted clips were not included during the task presentation. Instead, extracted clips were further cut to limit the degree of social information available and vary predictability for each interaction. The frame range (time period) shown for each clip was adjusted based on the "initiation" start frames classified in Blender. For the easy condition, modified clips began at the same start frame identified for the original extracted clips, and ended right at the "initiation" start frame or 200 ms before the "interaction" start frame. The hard condition started at the same time as the easy condition, but the clip ended 100 ms before the "initiation" start frame or 500 ms before the "interaction" start frame.

#### 4.3.4 Experimental Design

The task set consisted of 5 unique tasks. In every run, across behavioral and fMRI sessions, each task was performed 3 times for 35 sec (Fig. 2B). The 35 sec block was divided into a 5 sec instruction period, where the task name (for example, Visual Search task), the response effector ("Use your RIGHT hand") and the button-to-response assignment ("3 = item present, 4 = item absent") were presented on the screen. Participants were instructed to respond with their left hand for both Action Prediction (1 = right; 2 = left), and N Back (1 = Match; 2 = No Match), and they responded with their right hand for Semantic Prediction (3 = meaningful; 4 = meaningless) and Social Prediction (3 = hug; 4 = handshake). Task-specific instructions were followed by a 30 sec period of continuous task performance. Novel stimuli were introduced across the runs to prevent participants from learning specific stimulus-response associations. In the fMRI runs, each task block was followed by a 10 sec period of rest, during this time participants were instructed to relax and focus their attention on a fixation cross placed in the center of the screen. The motivation for testing all tasks within a scanning run, rather than testing one task in each run, was to ensure a common baseline across all tasks, enabling between-task comparisons. This structure was also replicated in the behavioral sessions to ensure consistency between session types. Four of the tasks

(visual search, N Back (picture), action prediction, social prediction) had a rapid, discrete trial structure (trials per block = 15), whereby each unique stimulus (letter, picture, video, video) was presented for 2 sec, with the response required to be completed within this window, followed by an intertrial interval (ITI) of 500 ms. The remaining task, semantic prediction, had a slower discrete trial structure (trials = 5, trial duration = 6 sec, ITI = 0s).

#### 4.3.5 Scanning Sessions

Participants completed three scanning sessions over the course of the 3-week experiment. In the first session (session 1), participants completed a short training block outside the scanner (10 trials per task) to familiarize them with the task instructions. Participants completed 6 11-min imaging runs, with a minimum 2 minute break between each run. The second and third scanning sessions each took place one week after the preceding session (mean number of days = 6.7; 7.4, s.d. = 2.5; 1.8). Sessions 2 and 3 were identical to the first scanning session, with the exception that participants did not complete an initial training block. Each scanning session consisted of six imaging runs (11 min per run). Each of the 5 tasks was presented three times for 35 sec in each imaging run, producing 54 independent measurements per task. Online feedback was not provided during the scanning sessions. This was done to avoid contamination of the BOLD signal with reward-related activation. However, overall feedback was provided at the end of each run. Participants were encouraged to respond as accurately and as quickly as possible. The task order was randomized across runs. To reduce order effects within each set, task blocks were not presented in the same order across runs. The order within each run, as well as the order of the runs, was kept constant for all participants. This procedure was chosen to allow for across-participant analyses on the time series level (results not presented here). As noted before, novel stimuli were used in each run to reduce the recall of specific stimulus-response associations.

#### 4.3.6 Behavioral Training

Participants completed two behavioral training sessions over the course of the 3-week experiment. The first training session (session 2) took place 2-3 days after the first fMRI session (mean number of days = 2.14, s.d. = .98) and the second session (session 4) similarly took place 2-3 days after the second fMRI session (mean number of days = 2.67, s.d. = 1.1). Participants completed 7 8-minute runs in each session, with a minimum 2 minute break between each run. Training included all of the tasks with the exception of the rest condition. Participants were instructed on the tasks, and when ready, they initiated the first run. We did not include a training block for the behavioral sessions given that participants had already completed all tasks in a previous scanning session. Unlike the scanning sessions, online feedback was provided for all tasks (green tick marks or red crosses to indicate correct or incorrect responses, respectively). At the end of each run, run-average accuracy and reaction time scores were provided to participants. Participants were encouraged to respond as accurately and as quickly as possible on each trial.

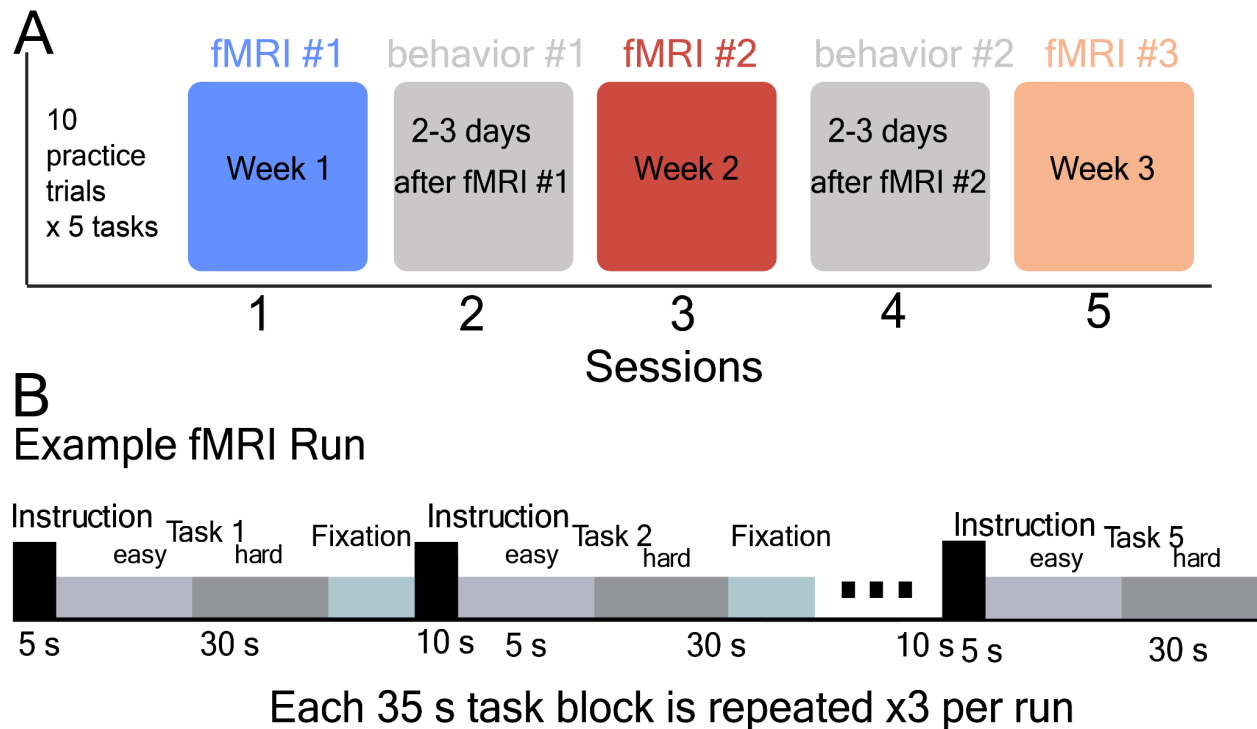


**Figure 2. Experiment Design.** A) Participants ( $n=25$ ) were trained on a battery of 5 cognitive tasks over the course of 3 weeks. There were 2 behavioral and 3 fMRI sessions, each of the fMRI sessions were separated by one week, and each of the behavioral sessions happened 2-3 days after the first and second fMRI sessions respectively. Participants completed 10 practice trials per task (~5 minutes total) before the first session so that they were familiar with the task instructions. B) Example functional run for fMRI sessions 1,3,5. Each task block was 35 sec long (5 sec instruction and 30 sec task execution). Each task was followed by 10 sec of rest (fixation cross). All 5 tasks were conducted 3 times per run (counterbalanced within each run), and each run was 665 sec in length. There were 6 runs per session.

#### 4.3.7 Image Acquisition

All fMRI and MRI data were collected on a SIEMENS MAGNETOM TrioTim syngo MR B17 in the Henry Wheeler Jr. Brain Imaging Center at the University of California, Berkeley. The protocol used the following parameters: 1 s repetition time; field-of-view measuring 20.8 cm; A-to-P phase encoding direction; 48 slices; 3 mm thickness; in-plane resolution  $2.5 \times 2.5$  mm<sup>2</sup>. In order to localize and normalize the functional data, a high-resolution anatomical scan (T1-weighted MPRAGE, 1mm isotropic resolution) of the whole brain was acquired during the first session.

#### 4.3.8 Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow

(from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, Zhang, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR\_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al. 2009; TemplateFlow ID: MNI152NLin2009cAsym).

#### 4.3.9 Functional data preprocessing

For each of the 18 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Susceptibility distortion correction (SDC) was omitted. The BOLD reference was then co-registered to the T1w reference using bregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al. 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox and Hyde 1997). The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsnative*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each *CompCor* decomposition, the  $k$

components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.6.1 (Abraham et al. 2014), mostly within the functional processing workflow. All preprocessing was conducted using *fMRIPrep* 20.0.0 (Esteban, Markiewicz, et al. 2018; Esteban, Blair, et al. 2018), which is based on *Nipype* 1.4.2 (Gorgolewski et al. 2011; Gorgolewski et al. 2018).

#### 4.3.10 General Linear Model (GLM)

To generate estimates of the activity related to each task condition, a General Linear Model (GLM) was fitted for each voxel to the time series data using *Nilearn* (Abraham et al. 2014). This was done separately for each imaging run with each session. Each task condition (including rest) was modeled as a separate regressor. The 5-second task-specific instructions were modeled as separate regressors in the GLM. We included instruction regressors for each of the five task conditions: N Back, Visual Search, Semantic Prediction, Action Prediction, and Social Prediction. The beta weights were averaged across all 6 runs for each session. The task set therefore was based on 48 activity estimates each (3 sessions x (5 Instructions + 11 condition regressors)).

#### 4.3.11 Cerebral Cortex Surface Reconstruction

For each of the participants, the anatomical surfaces of the cortical hemispheres were reconstructed using the standard recon-all pipeline of the freesurfer package (v. 5.0; Fischl, 2012). The pipeline included the following steps: brain extraction, generation of white and pial surfaces, inflation, and spherical alignment to the symmetric fsLR-32k template (Van Essen et al. 2012). Individual surfaces were re-sampled to this standard grid, resulting in surfaces with 32,492 vertices per hemisphere.

#### 4.3.12 Spatial Normalization of Cerebellar Data

The cerebellum was isolated using the Spatially Uninformed Infratentorial Template toolbox in SPM12 (SUIT v3.3; Diedrichsen, 2006) and then normalized to a high-resolution template of the human cerebellum using the Dartel algorithm (Ashburner, 2007). This non-linear transformation was applied to both the anatomical

and functional data. Task condition-evoked activity estimates (i.e. beta weights) were resampled to a resolution of 3 mm isotropic and resliced into SUIT space. The cerebellar isolation mask resulting from the isolation process was edited in order to exclude voxels in the superior cerebellum that abutted voxels in the primary visual cortex. Functional images were then masked with the cerebellar isolation mask in order to include activation signals originating from the cerebellar cortex alone. To visualize cerebellar data, we used a surface-based, flat-map representation of the cerebellar cortex provided by the SUIT toolbox (Diedrichsen & Zotow, 2015). The flat-map is intended only for display purposes as it does not represent a true unfolding of the cerebellar cortex, but rather an average of multiple folia.

### 4.3.13 Connectivity Models

We used connectivity weights from a pre-trained cortico-cerebellar model estimated on a large multi-domain task battery (MDTB) (see King et al. 2022 for details). **Ridge** regression was used to incorporate a model that penalizes large connectivity weights and therefore attempts to find a broad set of cortical regions to explain each cerebellar voxel. Ridge regression minimizes the following Loss:

$$\hat{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\mathbf{W}\|_2^2 + \lambda \|\mathbf{W}\|_2^2$$

The model was trained using 5 random and 5 functional parcellations of the neocortex , resulting in a total of 15 models (see King et al. 2022 for details on model estimation). Each of the models was trained on one set of tasks and then tested on a new dataset. In the current study, we used the pre-trained model that yielded the highest predictive accuracy on the MDTB data (L2-regularized model;  $\alpha = \log(8)$ ; 1848 cortical regions). For the purposes of the current study, we calculated the group weights of the pre-trained model, by averaging across participants. We then used these group weights to predict cerebellar activity patterns for the five unique tasks in our experiment. This prediction was done separately for each session, and for each participant.

### 4.3.14 Cortical parcels

The Q features of the models were defined by a parcellation of the cortex and by averaging condition-evoked activity estimates across these cortical regions. Cortical regions were created by subdividing each cortical hemisphere into a set of 1848 regions. Specifically, each hemisphere was divided into a set of hexagonal parcels, using a regular Icosahedron with 1002 parcels per hemisphere (see Zhi et al. 2021 for details). Parcels that were completely contained within the medial wall were excluded, resulting in 1848 parcels for the entire neocortex. The Icosahedron parcellation is arbitrary in the sense that it does not align to functional boundaries in the human neocortex more than expected by chance (Zhi et al. 2021). The analysis was therefore repeated with a different functionally derived parcellation from Schaefer et al. (2018). A publically available collection and functional evaluation of these cortical parcellations can be found in a recent paper by Zhi et al. (2021).

#### 4.3.15 Model Testing

Model testing was performed on the predicted time series, rather than on the basic activity estimates (beta weights) themselves. For this we premultiplied the activity estimates with the first-level design matrix. This procedure in effect reweights the activity estimates optimally, accounting for the fact that regressors representing instructions relied on 5s of fMRI data, while regressors representation conditions relied on 10-30s. This approach also addresses the problem of covariance of estimation errors in event-related designs (Mumford et al., 2012).

The predicted time series were normalized by subtracting the overall mean across conditions and by dividing them by the standard deviation of the residuals from the first-level GLM. The time series (of length  $T = 665$ ) were then averaged for all  $Q$  cortical parcels (see cortical parcels), resulting in a  $T \times Q$  matrix ( $\mathbf{X}$ ) of cortical features. For the cerebellum, we used all  $P=6937$  cerebellar voxels (SUIT atlas space with 3mm resolution), resulting in a  $T \times P$  matrix ( $\mathbf{Y}$ ).

#### 4.3.16 Reliability of cerebellar activity patterns

To determine intra-participant reliability across the entire cerebellar cortex, we calculated the correlation between the average activation estimates for odd and even runs for each session, separately for each participant. To obtain an overall reliability, we stacked the 16 (task conditions)  $\times$  3 (sessions) activation estimates for all cerebellar voxels into a single vector and calculated the Pearson correlation between the two estimates. The reliability was plotted for each session (see Supplementary Figure 6).

#### 4.3.17 Evaluating task-specific deviations from the model

We were interested in examining task-specific deviations from overall model prediction. Task-specific deviations from the connectivity model can be interpreted as unexplained variance in the observed activation for a task condition. Given that the connectivity weights are estimated on a dataset with different signal-to-noise ratio, we could not compare the predicted and observed activations directly for each of the tasks. Instead, we fit a linear regression model to relate the predicted and observed data. To meaningfully examine model residuals, we first standardized the observed and predicted data (as scale for each vector was different), and computed the model residuals. Data points falling above and below the regression line respectively, were interpreted as being under- and over predicted by the connectivity model.

### 4.4 Results

#### 4.4.1 Overview

We designed a multi-task learning experiment to assess whether changes in task-evoked cerebellar activation could be predicted by a model of cortico-cerebellar connectivity, across multiple stages of learning. Five tasks, each representing unique cognitive domains, were chosen for their ability to 1) be learned across multiple sessions, and to 2) activate both the cerebral cortex and the cerebellum in a widespread

fashion (King et al. 2019). There were two trial conditions in each task, easy and hard, to modulate task difficulty. Behavioral performance was measured across all five sessions (behavioral and scanning) to track participant learning. Accuracy was used as a metric of behavioral performance for two tasks: Action Prediction and Social Prediction, and reaction time was used to measure performance for the other three tasks: N Back, Visual Search, and Semantic Prediction.

We used model weights from the best task-independent model of cortico-cerebellar connectivity, previously determined by our group (King et al. 2022), to make predictions for cerebellar activity patterns across different stages of learning. The model that we used had been trained to predict activity in each voxel of the human cerebellum as a linear function of cortical activity patterns (see King et al. 2022 for details), using cortical and cerebellar fMRI data obtained from 24 participants who performed a large task battery over four scanning sessions. The weights of this model were estimated with Ridge Regression (L2 regularization), which allowed for a wider distribution of cortical inputs onto cerebellar voxels. Models were trained and validated with a different number of cortical parcellations, varying in their degree of granularity. Here, we evaluated the model, which had been estimated with the largest number of cortical regions (1848 regions). Weights  $\hat{W}$  from the model were then used to make predictions for each task-evoked cerebellar activity pattern across sessions 1, 3, and 5 of the multi-task learning dataset (Fig. 2A).

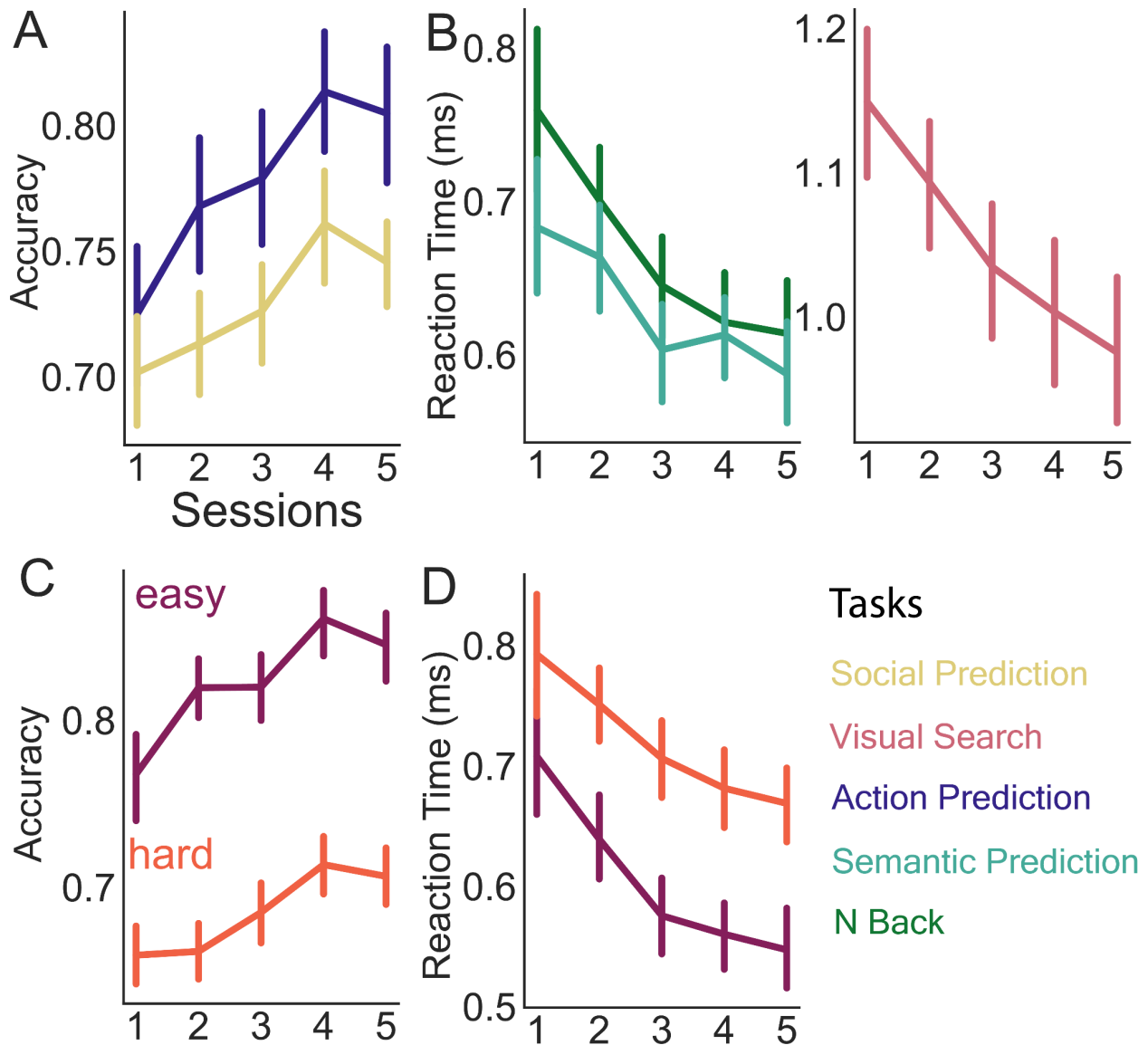


**Figure 3. Behavioral performance improves across sessions.** A) Accuracy improves in both the Action Prediction and Social Prediction tasks, B) Reaction time decreases for N Back, Semantic Prediction, and Visual Search tasks, C) Accuracy increases for both easy and hard conditions (averaged across Social Prediction and Action Prediction tasks), and D) Reaction time decreases for both easy and hard conditions (averaged across Visual Search, N Back, and Semantic Prediction tasks). Sessions 1, 3, 5 are scanning sessions; sessions 2 and 4 are behavioral sessions.

#### 4.4.2 Participants learn cognitive tasks across multiple weeks

We trained 22 participants to perform 5 2AFC cognitive tasks, each representing a unique domain: perception, social cognition, action observation, working memory/executive function, and language (see Supplementary Table 1 for details). During behavioral training (sessions 2 and 4), participants received visual feedback on each trial (green for correct and red for incorrect responses), and they were encouraged

to respond as quickly and as accurately as possible (see Materials and Methods). During fMRI testing (sessions 1, 3, 5), participants did not receive trial-by-trial feedback, instead, feedback was displayed at the end of every run. Accuracy was used as a performance metric for the Action Prediction and Social Prediction tasks, and reaction time was used to measure performance for the N Back, Visual Search, and Semantic Prediction tasks. For the latter 3 tasks, accuracy was stable and close to ceiling performance (>80%) across all 5 sessions (Supplementary Figure 7), mitigating any effects of speed-accuracy tradeoffs.

An ANOVA revealed a significant increase in accuracy across the 5 sessions for Action Prediction and Social Prediction tasks (Fig 3A;  $F_{4,92} = 33.59$ ,  $p < .001$ ,  $\eta^2 = 0.59$ ), and a significant decrease in reaction time for the Visual Search, N Back, and Semantic Prediction tasks (Fig 3B;  $F_{4,92} = 70.07$ ,  $p < .001$ ,  $\eta^2 = .75$ ). There was a significant main effect of task difficulty (easy and hard) for both accuracy ( $F_{1,23} = 809.03$ ,  $p < .001$ ,  $\eta^2 = .97$ ) and reaction time measures of behavioral performance ( $F_{1,23} = 424.57$ ,  $p < .001$ ,  $\eta^2 = 0.94$ ). Post hoc t-tests, corrected for multiple comparisons, revealed a significant difference in accuracy across the behavioral sessions (sessions 2 and 4;  $t_{23} = -6.39$ ,  $p_{\text{corr}} < .001$ ), and the first and last fMRI sessions (sessions 1 and 5;  $t_{23} = -7.46$ ,  $p_{\text{corr}} < .001$ ). Similarly, FDR-corrected post hoc t-tests revealed a significant difference in reaction time across the behavioral sessions ( $t_{23} = 8.246$ ,  $p_{\text{corr}} < .001$ ) and the first and last fMRI sessions ( $t_{23} = 11.91$ ,  $p_{\text{corr}} < .001$ ), with reaction time decreasing across sessions, indicating that participants successfully learned the tasks. Additionally, the interaction between sessions and task difficulty was significant for both accuracy (Fig 3C;  $F_{4,92} = 6.95$ ,  $p < .001$ ,  $\eta^2 = 0.23$ ), and reaction time (Fig 3D;  $F_{4,92} = 16.44$ ,  $p < .001$ ,  $\eta^2 = .42$ ), and FDR-corrected post hoc tests revealed that this interaction was significant across all sessions. Thus, participants demonstrated, through improvements in behavioral performance, that they were able to successfully learn the tasks across three weeks of training (see Supplementary Figure 8 for performance across runs). Moreover, while participants were able to learn both easy and hard conditions, there were significant differences in overall accuracy and reaction time performance across levels of difficulty. These behavioral results now allow us to identify learning-related changes in blood oxygen-level dependent (BOLD) activation patterns for each of our five tasks, and across levels of difficulty.

#### 4.4.3 Cortico-Cerebellar Connectivity Model Predicts New Tasks

We used weights from a cortico-cerebellar connectivity model pre-trained with L2 regularized regression (1848 cortical regions) to make predictions for cerebellar activity patterns of five unique tasks. Each of the tasks were learned over the course of 5 sessions, including 2 behavioral and 3 scanning sessions. We used the pre-trained weights to make predictions separately for each scanning session, and then averaged the observed data and predictions across all three sessions to determine overall task deviations from the connectivity model.

We previously reported that pre-trained, L2-regularized connectivity models, could reasonably well predict the variation of cerebellar activity across a new taskset (King et al. 2022). Given this gap in model prediction, we were interested in examining overall task-specific deviations using a unique set of cognitive tasks. Deviations from

the connectivity model can be interpreted as unexplained variance in the observed activation for a particular task condition. Given that the connectivity weights are estimated on a dataset with different signal-to-noise ratio, we could not compare the predicted and observed activations directly for each of the tasks. Instead, we fit a linear regression to relate the predicted and observed data (Fig 4A). While the overall prediction was strong ( $R=63\%$ ), there were significant task deviations from the regression line. We were interested in knowing which tasks fell “above” and “below” the line. Tasks that fall above the line are underestimated by the connectivity model, which means that the linear model is not complex enough to capture the variance in the cerebellar data for a given task. A task that falls below the line is overestimated by the connectivity model, in this case the model itself is too complex for the observed task-evoked activity patterns in the cerebellum. Tasks that fall on the regression line are perfectly predicted by the model.

First, we wanted to examine differences in overall activity patterns across the tasks and their respective conditions (easy, hard, instructions). For each of the task conditions, standard error of cerebellar activity patterns (vertical line), and standard error of model predictions (horizontal line) are plotted (Fig 4A). Tasks that activate the cerebellum most strongly are Social Prediction and Action Prediction. For each of the tasks, the “hard” task condition activates the cerebellum more than the “easy” condition, while task-specific instructions weakly activate the cerebellum (Fig 4A).

Next, we examined model residuals to identify task-specific deviations from the regression line (Fig 4B). In order to meaningfully interpret model deviations, we first standardized the observed data and the predictions, and then computed the residuals. A 2-way ANOVA between Task and Trial Type (easy, hard, instructions) revealed a main effect of Task ( $F_{4,315} = 3.247$ ,  $p < .01$ ) and Trial Type ( $F_{2,315} = 42.91$ ,  $p < .001$ ), and there was a significant interaction between both factors ( $F_{8,315} = 3.84$ ,  $p < .001$ ). Post-hoc tests (FDR-corrected) found a main effect between the instructions and hard ( $t_{218} = 3.41$ ,  $p_{\text{corr}} < .001$ ) conditions, and the instructions and easy ( $t_{218} = 4.87$ ,  $p_{\text{corr}} < .001$ ) conditions. However, there was no main effect for easy and hard conditions ( $t_{218} = 2.6$ ,  $p_{\text{corr}} = .11$ ). The only exception was the N-Back task, a two-sided t-test revealed a significant difference between easy and hard conditions ( $t_{(21)} = 2.46$ ,  $p_{\text{corr}} < .05$ ). Despite the fact that there were significant behavioral differences between easy and hard conditions across all tasks, there is no evidence that a model of cortico-cerebellar connectivity distinguishes between levels of task difficulty in predicting cerebellar activity. The only exception to this result is the N Back task in which the hard and easy conditions are under- and over predicted, respectively.

Finally, we examined the model residuals for each of the tasks without the influence of the instruction period (Fig 4C). We found that two tasks, Action Prediction and Social Prediction, were underpredicted by the model, while Visual Search and N Back were over predicted. Semantic Prediction was well predicted by the model, and a t-test did not find a significant difference between model residuals and zero ( $t_{21} = 0.84$ ,  $p = .56$ ). These results suggest that activity patterns evoked by Action Prediction and Social Prediction contain variance resulting from a nonlinear increase in activation, that is unexplained by a linear model of connectivity. However, we infer the opposite from Visual Search, namely, the connectivity model is too complex to accurately capture the variance of activity patterns evoked by this task, resulting in an overprediction of the

cerebellar data. Therefore, while the cortico-cerebellar connectivity model is successful overall at predicting novel cerebellar activity patterns, there are characteristic deviations of this linear model that suggest task-specific contributions of the cerebellum that cannot be explained by commensurate cortical activation.

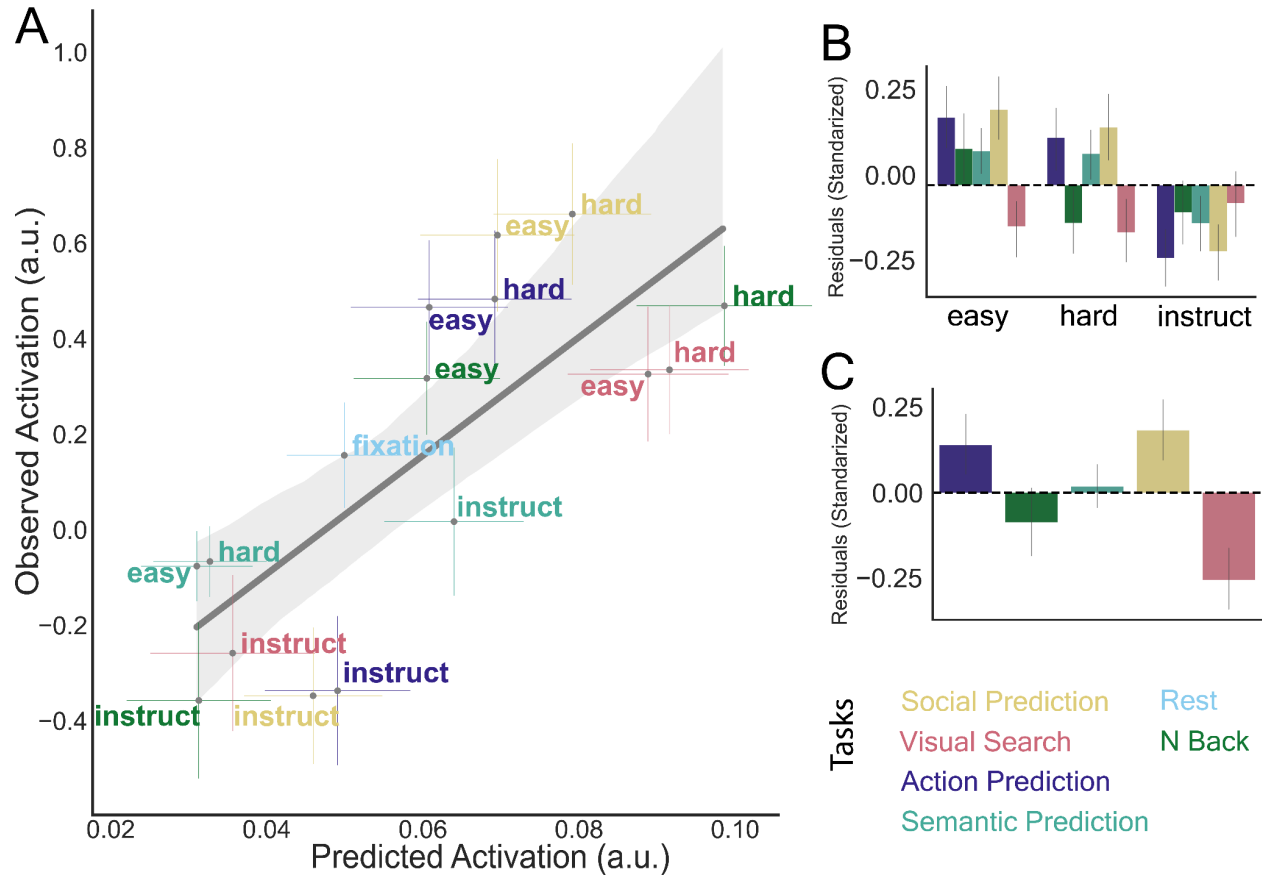


**Figure 4. Task-specific deviations from linear model of cortico-cerebellar connectivity.** A) A connectivity model estimated on cortical activation predicts cerebellar activation reasonably well, although there are characteristic violations. B) Residuals computed on standardized observed and predicted data. Instructions are overpredicted by the model for each task, and there is no significant difference in prediction for easy and hard conditions. C) Action Prediction and Social Prediction evoke high cerebellar activity that is underpredicted by the model, while Visual Search is overpredicted by the model.

#### 4.4.4 Cerebellar BOLD signal decreases across learning

In line with traditional fMRI analysis, we examined task-specific changes in cerebellar activity patterns (i.e. overall BOLD activation) across the sessions (Fig. 5). A 2-way ANOVA on the observed cerebellar activity patterns revealed a main effect of Session ( $F_{4,84} = 26.35$ ,  $p < .001$ ) and Task ( $F_{2,42} = 5.0847$ ,  $p < .01$ ). There was also a significant interaction between both variables ( $F_{8,168} = 2.01$ ,  $p < .05$ ). Post-hoc tests, FDR-corrected for multiple comparisons, found that there was no significant difference between activity patterns across sessions 1 and 3 ( $t_{21} = 0.35$ ,  $p = .97$ ) but there was a significant difference across sessions 3 and 5 ( $t_{21} = 5.01$ ,  $p < .001$ ) and sessions 1 and 5 ( $t_{21} =$

4.88,  $p < .001$ ). Overall, we observed that cerebellar activity decreased across tasks between the first and last scanning session, with the largest decrease occurring between sessions 3 and 5 for all tasks.

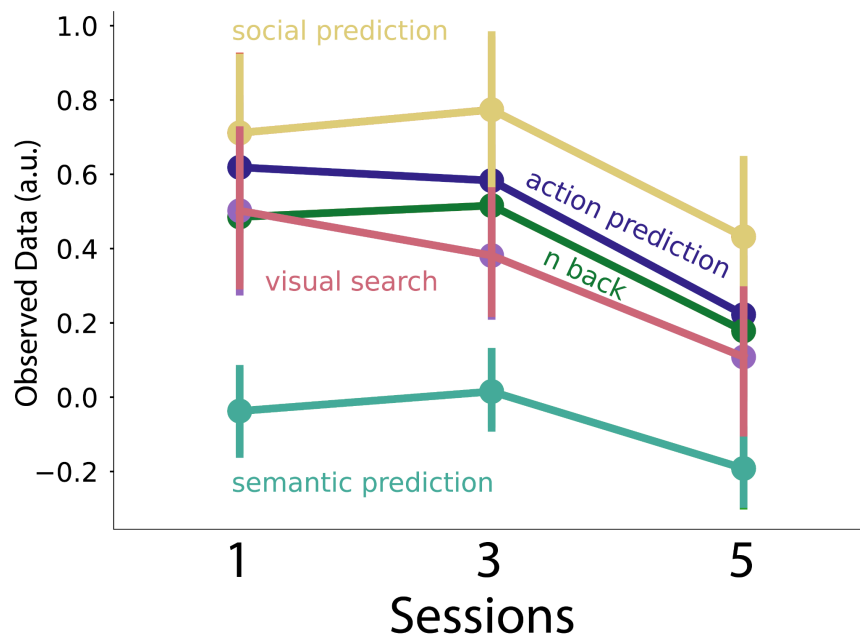


**Figure 5. Decrease in cerebellar BOLD signal (i.e., beta weights) across scanning sessions for each task.**

#### 4.4.5 Early learning is underpredicted by a model of cortico-cerebellar connectivity, but only for dynamic prediction tasks

However, it is not enough to know that cerebellar activity patterns decrease across sessions. Instead, we wanted to determine the extent to which changes in cerebellar BOLD across learning could be predicted by commensurate changes in cortical activity patterns. Specifically, we wanted to determine whether there would be a difference in model predictions for cerebellar activity early in learning (session 1) compared to predictions late in learning (session 5). We hypothesized that if the cerebellum's role is to facilitate cortico-cortico communication in learning new representations, then cerebellar activity should be highest during early learning. A linear connectivity model would underpredict cerebellar activity early in learning (i.e., tasks would be above the line for session 1), and once the cerebellum “drops out” of learning, the model should be better able to capture variance in cerebellar activity patterns. As before, we used weights from a cortico-cerebellar connectivity model pre-trained with L2 regularized regression (1848 cortical regions) to make predictions for cerebellar activity patterns of the five unique tasks. We used the weights to make model predictions, separately for each scanning session, and then fit a linear regression to relate the observed and predicted activity patterns. We opted not to include task-specific instruction periods or the rest period in the main analysis as we were more interested in knowing how the model predicted changes related to learning rather than task instruction and fixation.

We examined model residuals to identify whether there was a significant difference in task-specific deviations from the overall regression line, separately for each session (Fig. 6). A 2-way ANOVA between Session and Task on the model residuals revealed a main effect of Session ( $F_{2,42} = 7.21$ ,  $p < .01$ ), and Task ( $F_{1,21} = 41.01$ ,  $p < .001$ ), although the interaction between these two factors was not significant ( $F_{2,42} = 1.91$ ,  $p = .06$ ). Post-hoc t-tests (FDR-corrected) revealed that the main effect of Session was significant across sessions 3 and 5 ( $t_{21} = 2.4$ ,  $p = .03$ ), and sessions 1 and 5 ( $t_{21} = 3.32$ ,  $p < .01$ ), however, there was no significant difference in the model residuals between sessions 1 and 3 ( $t_{21} = -.207$ ,  $p = .84$ ). What we can interpret from these results is that the connectivity model makes a similar prediction for cerebellar activity patterns in the first two sessions, but a significantly different prediction in the final session. This result likely reflects the fact that cerebellar BOLD is decreasing across sessions and by the final session, learning-related variance in cerebellar activity is better predicted by cortical activations, compared to earlier sessions.

We also examined model residuals to determine whether predictions were modulated by task difficulty (easy and hard conditions). While we did not find a main effect of Difficulty ( $F_{1,21} = 3.03$ ,  $p = .09$ ), we found a significant interaction between Session and Difficulty ( $F_{2,42} = 7.145$ ,  $p < .01$ ), and FDR-corrected post hoc t-tests revealed a significant interaction between Session and Difficulty for both sessions 3 ( $t_{21} = -3.25$ ,  $p < .01$ ) and 5 ( $t_{21} = -2.31$ ,  $p < .05$ ) but not session 1 ( $t_{21} = 1.32$ ,  $p = .99$ ). In both sessions 3 and 5, model residuals were greater for hard compared to easy conditions, suggesting that there is more unexplained variance in the cerebellum for difficult task conditions in later sessions.

Finally, there were some interesting patterns in learning-related changes across task conditions. For example, the only two tasks that were always underpredicted by the connectivity model were the Action Prediction and Social Prediction tasks (Fig. 6). Although the model was performing better by the final session, the model never achieved a perfection prediction, suggesting that there is unexplained variance at each stage of learning that cannot be captured by commensurate activation in the cerebral cortex. The opposite pattern was true for Visual Search: this task was overpredicted by the model for each session (Fig. 6). This means that the model itself is more complex than the data that are being predicted. One explanation for this result is that the cortical activity patterns used to train the connectivity model capture a lot of visual information, introducing a bias to overpredict tasks, such as Visual Search, that traditionally activate “visual” regions of the cerebellum (King et al. 2019). The remaining tasks, Semantic Prediction and N Back (Supplementary Fig. 9), were characterized by different patterns of model prediction. For example, Semantic Prediction was overpredicted by the connectivity model during the first session but underpredicted in sessions 3 and 5 (Fig. 6). It is worth noting that predictions for this task were closest to the regression line compared to the other tasks, suggesting that a cortico-cerebellar connectivity model is reasonably sufficient to account for variance in cerebellar activity evoked by Semantic Prediction. Interestingly, the N Back task displayed a unique pattern of prediction, such that easy task conditions were always underpredicted by the model, while hard conditions (with the exception of Session 2), were overpredicted (Supplementary Fig.

9). This result suggests that these two levels of the N Back task engage the cerebellum in different computations across stages of learning. Overall, we conclude from these results that two tasks in particular, Action Prediction and Social Prediction, evoke cerebellar activity patterns early in learning that cannot be predicted by cortical activity.

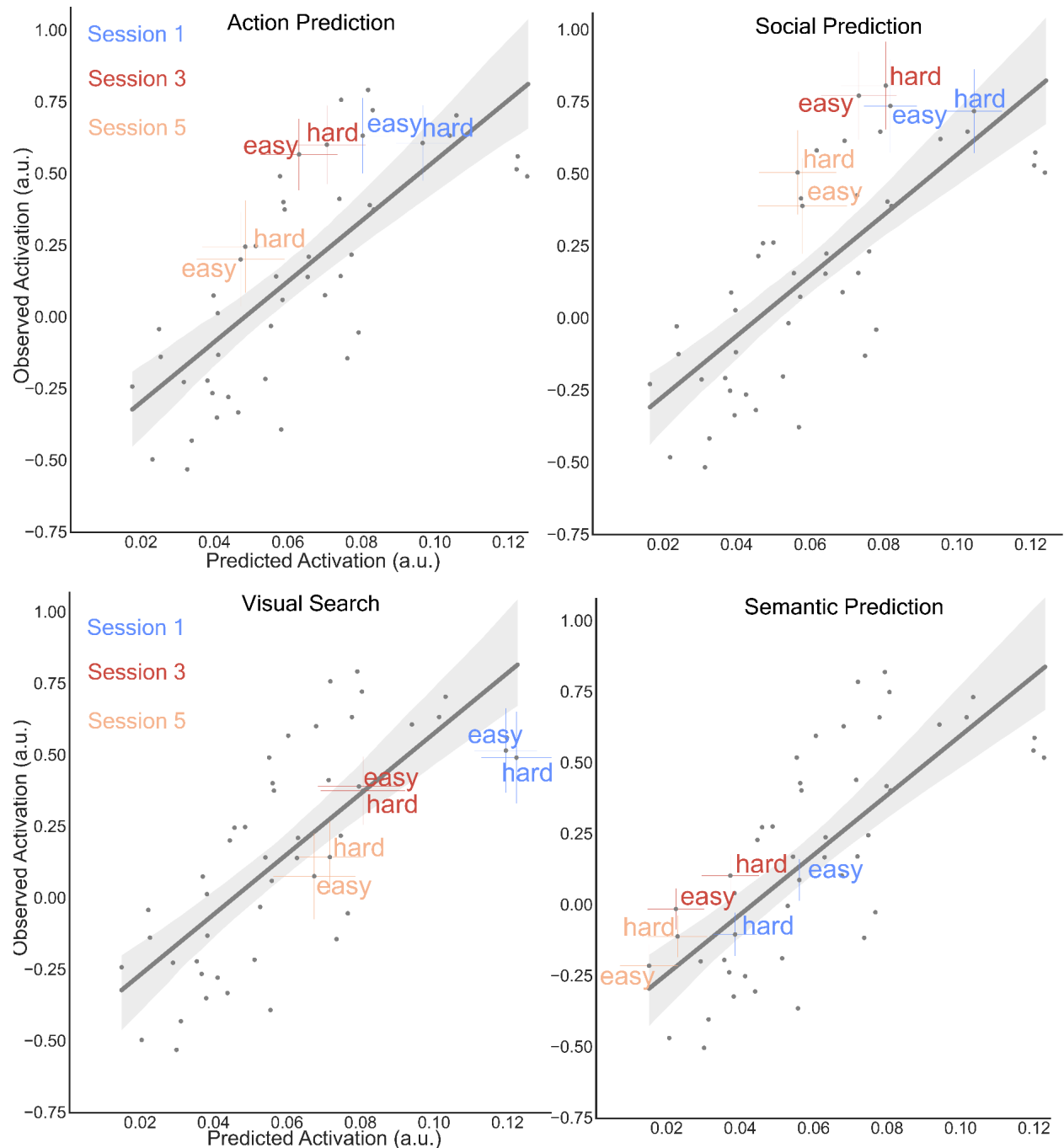


**Figure 6. Cortico-cerebellar connectivity model underpredicts cerebellar activity early in learning for dynamic prediction tasks.** *Dynamic Prediction tasks: Action Prediction and Social Prediction, are underpredicted by a linear connectivity model early in learning (Session 1), while Visual Search and Semantic Prediction are overpredicted*

*by the model. Note that the linear regression line is estimated for all 5 tasks (including easy and hard conditions), but for the purposes of visualization (e.g., to avoid overcrowding), each task, and their respective conditions, are shown separately.*

## 4.5 Discussion

### 4.5.1 Summary

Traditional neuroimaging approaches typically infer cerebellar function without considering relative contributions from task-evoked activity in the cerebral cortex. However, we know that the BOLD signal measured in the cerebellum likely reflects inputs from the cerebral cortex via mossy fiber projections (Mathiesen et al. 2000). Therefore, it is likely that some of the task-evoked activity patterns observed in the cerebellum mirror contributions from corresponding regions of the cerebral cortex. Here, we set out to investigate unique cerebellar contributions to cognitive function, accounting for commensurate changes in cortical activity patterns. To do this, we evaluated a pre-trained linear model of cortico-cerebellar connectivity on 5 unique cognitive tasks across three separate stages of learning. We measured task-specific violations of the linear model by observing the relationship between the observed and predicted cerebellar data. We know from previous work that this pre-trained model, estimated with cortical activity patterns alone, could reasonably predict novel cerebellar activity patterns (King et al. 2022), which suggests that each area of the cerebellum simply receives, independent of task, a copy of the neural activity occurring in the corresponding cortical areas. Importantly, the prediction performance of this model was not perfect, allowing for the model to fail in a number of characteristic ways. We found task-specific violations of the model, namely tasks that rely on dynamic action prediction processes: Action Prediction and Social Prediction, were underestimated by the model early in learning, while a task that strongly activates visual processes (Visual Search), was overestimated. We found tentative support for the early learning hypothesis: namely, the model underpredicted cerebellar activity early in learning. However, this hypothesis only held true for tasks that engaged prediction-based learning in dynamic action, and not semantic processing. We conclude that the cerebellum might play a task-specific role in supporting cortical processes early in the learning process when novel representations are first being acquired by the cerebral cortex.

### 4.5.2 Behavioral changes across learning

In this study, we evaluated learning-related changes in behavioral performance across 5 sessions (2 behavioral, 3 scanning). Reaction time was used as a metric of behavioral performance for Visual Search, Semantic Prediction, and N Back, and accuracy was used to measure changes in performance for Action Prediction and Social Prediction. We observed that performance improved across the sessions for each task (and task condition), corroborating previous literature. For example, improvement on the N Back working memory task has been observed across multiple days of training (Stojanaoski et al. 2015), while in a neuropsychological study of Semantic Prediction, we observe a similar decrease in reaction time across task blocks (King et al. 2022). Further,

decrease in reaction time across runs of the Visual Search task has been attributed to an improved search strategy (Sireteanu & Rettenbach, 1995). Reaction time differences between task conditions have also been demonstrated in previous work. For example, slower reaction times have been reported for violated compared to true predictions in a semantic verification task (D'Mello et al. 2017; Moberget et al. 2014), while reaction time has been shown to be modulated for set size in a Visual Search task (Sireteanu & Rettenbach, 1995). We also introduced two new behavioral tasks in this experiment: Action Prediction and Social Prediction. Both of these tasks engage dynamic prediction-based processes, and we found that participants could learn to accurately predict the correct outcome for both tasks. Task difficulty was modulated for both of these tasks, and participants were less accurate in predicting hard compared to easy task conditions. Overall, we demonstrated that participants improved their behavioral performance across all sessions for each cognitive task.

### 4.5.3 Unique cerebellar contribution to dynamic action prediction tasks

Previous neuroimaging research has reported cerebellar activation in response to a large number of cognitive tasks, including tasks employed in the current study: Visual Search (Schulman et al. 1997; King et al. 2019), N Back (Stoodley et al. 2012; King et al. 2019), and Semantic Prediction (Moberget et al. 2014; King et al. 2019). Although neither the Action Prediction nor Social Prediction tasks were tested in previous studies, similar tasks have evoked cerebellar activation in response to action observation (King et al. 2019; Errante & Fogassi, 2020) and social cognition (Van Overwalle et al. 2020). Furthermore, both of these tasks were intentionally presented as short video clips because we know from previous work that the cerebellum is strongly activated by dynamic processes (Nguyen et al. 2017; Chan et al. 2019; King et al. 2019).

One puzzle presented by neuroimaging research is that while the cerebellum is activated in response to many cognitive tasks, it is not at all clear what this structure is uniquely contributing to these task processes. One of the reasons for this ambiguity is that cerebellar activation is often analyzed and interpreted in isolation of corresponding cortical activation. This approach is likely problematic as we have known for a long time that a majority of the cerebral cortex projects parasynaptically, via the pontine nuclei, to distinct functional subdivisions of the cerebellar cortex (Kelly & Strick, 2003). Therefore, it is possible that task-evoked cerebellar activation is mirroring activity patterns in corresponding regions of the cerebral cortex. To account for potential cortical contributions to cerebellar activity patterns, we tested whether a model trained only on task-evoked cortical activity patterns, could predict cerebellar activation in a new set of cognitive tasks. We found that tasks related to dynamic prediction (e.g., Action Prediction and Social Prediction) were underpredicted by the connectivity model, while tasks related to working memory (N Back), and perception (Visual Search) were overpredicted by the model. Although the Semantic Prediction task was slightly underpredicted by the model, task-specific model residuals were not significantly different from zero, suggesting that this task was well predicted by cortical activity patterns.

We interpret underprediction of the model as an inability for cortical activity patterns to explain cerebellar variance for tasks that engage dynamic prediction processes. Conversely, overprediction of task activity patterns indicates that the model

is too complex for the data being predicted, namely, tasks related to perceptual processing and working memory. In many ways, it is not entirely surprising that the tasks underpredicted by the connectivity model are dynamic prediction-based tasks. First, classic theories of cerebellar function posit a critical role of the cerebellum in a wide range of predictive processes, from sensory error-based prediction (Schlerf et al. 2012), to adaptive (Sokolov et al. 2017) and action prediction (Blakemore & Sirigu, 2003). Moreover, both of these tasks necessitate that participants make a prediction about an upcoming action, for example: left or right kick for the Action Prediction task and a hug or handshake for the Social Prediction task. Therefore, in line with the internal model hypothesis (Ito, 2008), it is possible that the cerebellum is simulating patterns of spatially structured actions in order to generate an internal model that can then be used to make predictions about upcoming events (Leggio & Molinari, 2015).

#### 4.5.4 Limited evidence that the cerebellum plays a unique role in semantic prediction

The internal model hypothesis has also been proposed to explain cerebellar activation in response to a Semantic Prediction task (Moberget et al. 2014; D'Mello et al. 2017). However, in the current work, we do not find evidence for a unique cerebellar role in predicting the outcome of a semantic sentence. Instead, we propose that the cerebellum uniquely engages tasks that require a dynamic, or action-based, prediction. Indeed, there is a lot of evidence to suggest that the cerebellum is strongly engaged by action (Blakemore & Sirigu, 2003), and social based sequences (Heleven et al. 2021; Van Overwalle et al. 2019). We can interpret these results in the context of embodied cognition and adaptive control (Barton, 2012), suggesting that the cerebellum plays a critical role in the process of grounding knowledge about cognition in the sensorimotor system, thereby linking movement to thought (Guell et al. 2018). It is possible that we do not observe a unique cerebellar role for Semantic Prediction since this task engages predictions related to semantic content, rather than action cognition. Further support for the current result comes from the misalignment between the neuroimaging and neuropsychology literatures. Namely, fMRI research has historically implicated the cerebellum in semantic processing, however, patient work provides a mixed bag of results, with many studies, including recent work from our group (King et al. 2022), demonstrating that patients with cerebellar degeneration are not impaired on language-based tasks such as Semantic Prediction.

#### 4.5.5 Cerebellum plays a task-specific role in coordinating cortico-cortico interactions during early learning

We have demonstrated that the cerebellum may play a unique functional role in processing dynamic action-based predictions. However, we were also interested in knowing whether the cerebellum plays a broader mechanistic role in cognition, namely in coordinating cortico-cortico interactions to ensure smooth coordination in processes related to both action and cognition. Indeed, previous work has proposed that the vast pattern recognition capacity of the cerebellum could work in tandem with the cerebral cortex to refine or tune cortical operations (Sokolov et al. 2017; Wang et al. 2014), without necessarily generating a unique computational function. In the current study, we

were interested in assessing learning-related BOLD activity in the cerebellum while participants acquired new representations across cognitive tasks.

We first demonstrated that observed cerebellar activity decreased across sessions for all tasks. Previous literature has reported mixed results with regard to changes in BOLD activation across learning. For example, a number of neuroimaging studies have demonstrated that cerebellar BOLD signal increases during learning (Imamizu et al. 2000; Grafton et al. 2003), while another set of studies have reported the inverse (Miall & Jenkinson, 2005; Jupetner et al. 1997). However, to our knowledge, none of these studies accounted for changes in commensurate cortical activity patterns. In the current work, we evaluated how well a model, trained on cortical activations alone, could predict novel cerebellar activity across learning. When we compared observed and predicted cerebellar activity, we found that in session 1, only two tasks: Action Prediction and Social Prediction, were strongly underpredicted by the connectivity model. These results suggest that during early stages of learning, there is a flow of information between the cerebral cortex and the cerebellum, gated by nonlinear integration in the pontine nuclei (Thier & Schwarz, 1999). We suggest that since the connectivity model only captures a linear relationship between the cerebral cortex and cerebellum, it fails to account for the large amount of nonlinear information arriving at the cerebellar cortex early in learning, thus resulting in an underprediction of the cerebellar data. We found the opposite pattern for the Visual Search and Semantic Prediction tasks, namely, the model overpredicted cerebellar data during early learning (session 1), which, if our hypothesis of early learning is correct, suggests that the cerebellum may not have been as engaged during early learning for these tasks. Importantly, while we found that observed cerebellar BOLD activity decreased across learning for all tasks, only two tasks: Action Prediction and Social Prediction, were underpredicted by the connectivity model during early learning. Based on these findings, we tentatively suggest that the cerebellum plays a task-specific role in consolidating cortico-cortico interactions during early learning.

#### 4.5.6 Conclusions

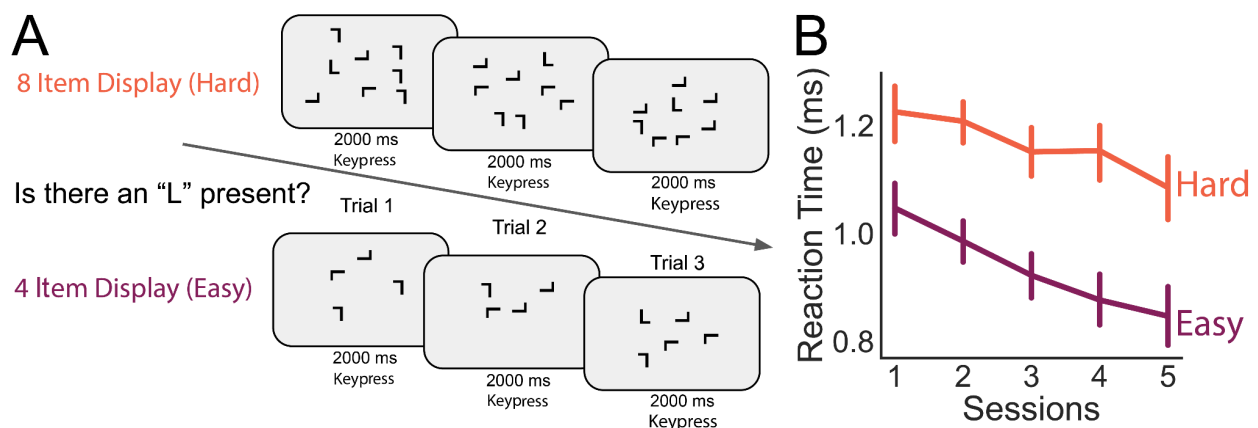
Our approach was designed to test functional and mechanistic theories of the cerebellum using a novel neuroimaging approach. We leveraged a general model of cortico-cerebellar connectivity, which used cortical activity patterns to make predictions of cerebellar activity in a set of 5 unique tasks, across different stages of learning. While the model provided a reasonable prediction of cerebellar activity, we were particularly interested in examining task-dependent deviations from the model prediction, allowing us to make important insights about cerebellar function. We found that certain tasks, namely those that engaged dynamic action prediction processes, activated the cerebellum more than predicted by the commensurate cortical activity. Further, these tasks were strongly underpredicted by the model early in the learning process compared to tasks engaging perceptual and working memory processes. This pattern of results suggests a task-dependent “gating” of cortical input to the cerebellum, likely within the pontine nuclei, with the strongest flow of information occurring early in learning. Therefore, we suggest that rather than transmitting a copy of cortical input to the cerebellar cortex, information arriving from the neocortex is modulated as a function

of the relative importance of cerebellar processing for tasks engaging specific cerebellar-dependent processes, such as dynamic action prediction.

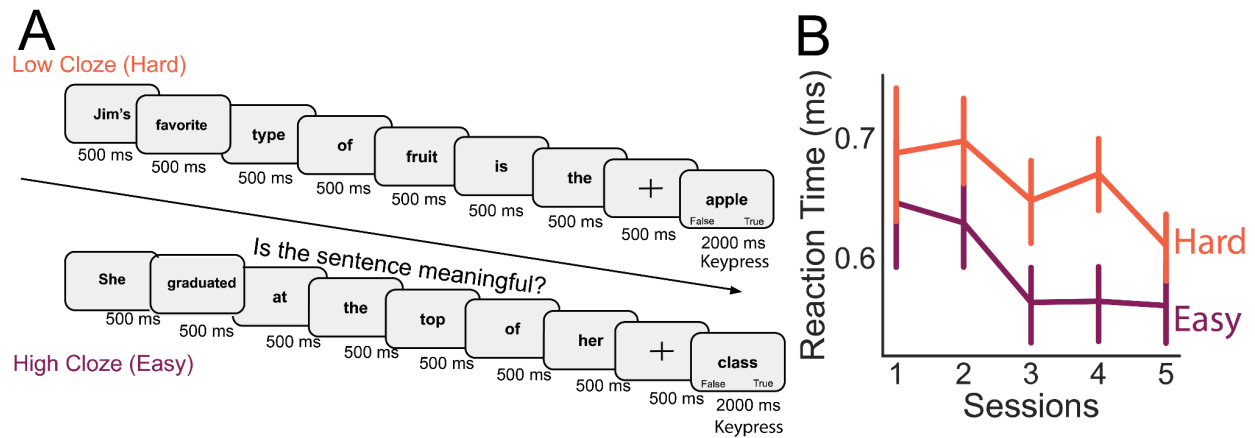
## 4.6 Supplementary Figures

| Domain             | Task Name           | Task Description                                                                                                   | Condition                             |
|--------------------|---------------------|--------------------------------------------------------------------------------------------------------------------|---------------------------------------|
| Action Observation | Action Prediction   | 2AFC task to predict if soccer players in a video clip will kick the ball to the left or right of the goalkeeper.  | Goal (easy)<br>Miss (hard)            |
| Perception         | Visual Search       | 2AFC, indicating if target stimulus ("L") is present among distractors (rotated "L") with varying set size (4, 8). | 4 item (easy)<br>8 item (hard)        |
| Language           | Semantic Prediction | 2AFC task to indicate if sequentially-presented words comprise a semantically meaningful sentence.                 | High Cloze (easy)<br>Low Cloze (hard) |
| Working Memory     | N Back              | 2AFC, indicating if current stimulus in stream of letters matches letter displayed two items previously (2-back).  | N Back+ (easy)<br>N Back- (hard)      |
| Social Cognition   | Social Prediction   | 2AFC task to predict if actors in a video clip will shake hands or hug at the end of the social interaction.       | 0ms (easy)<br>-120ms (hard)           |

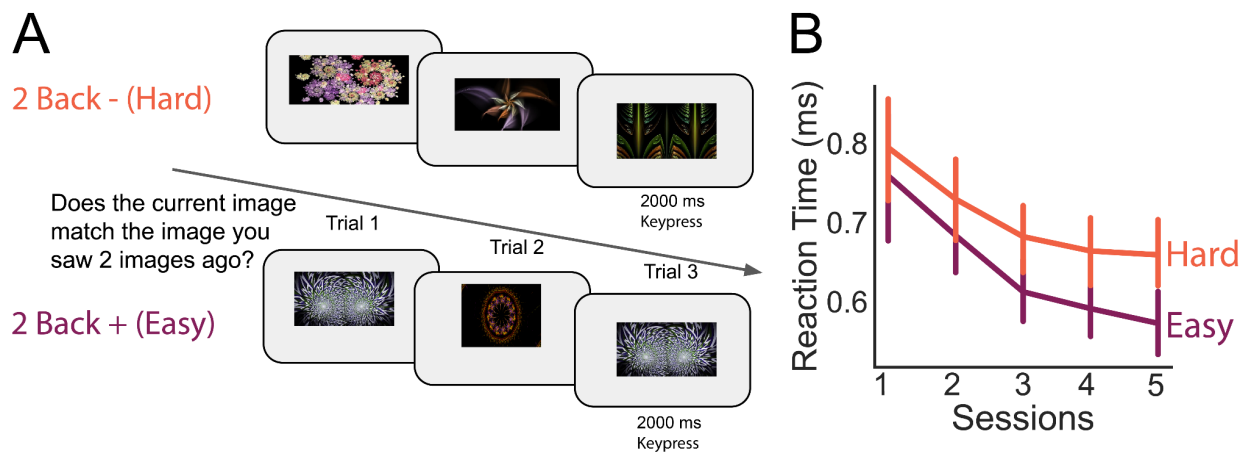
**Table 1. Task Description.** Task set and domain description for all 5 unique 2AFC tasks and 10 unique conditions, modulated across difficulty levels (easy and hard).



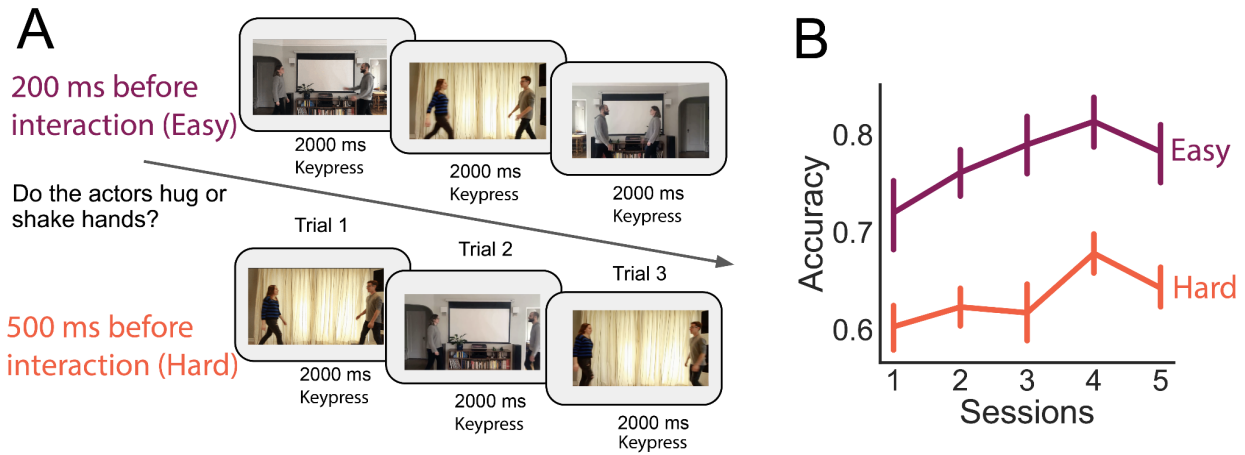
**Supplementary Figure 1. Task design for visual search.** A) Two alternative forced choice task (2AFC), in which participants indicate "True" if the target stimulus "L" is present among distractor stimuli (rotated "L") or "False" if it is not present. The set size varies to modulate difficulty, there are 4 items for the easy condition, and 8 items for the hard condition. B) Reaction time performance across sessions for both easy and hard conditions.



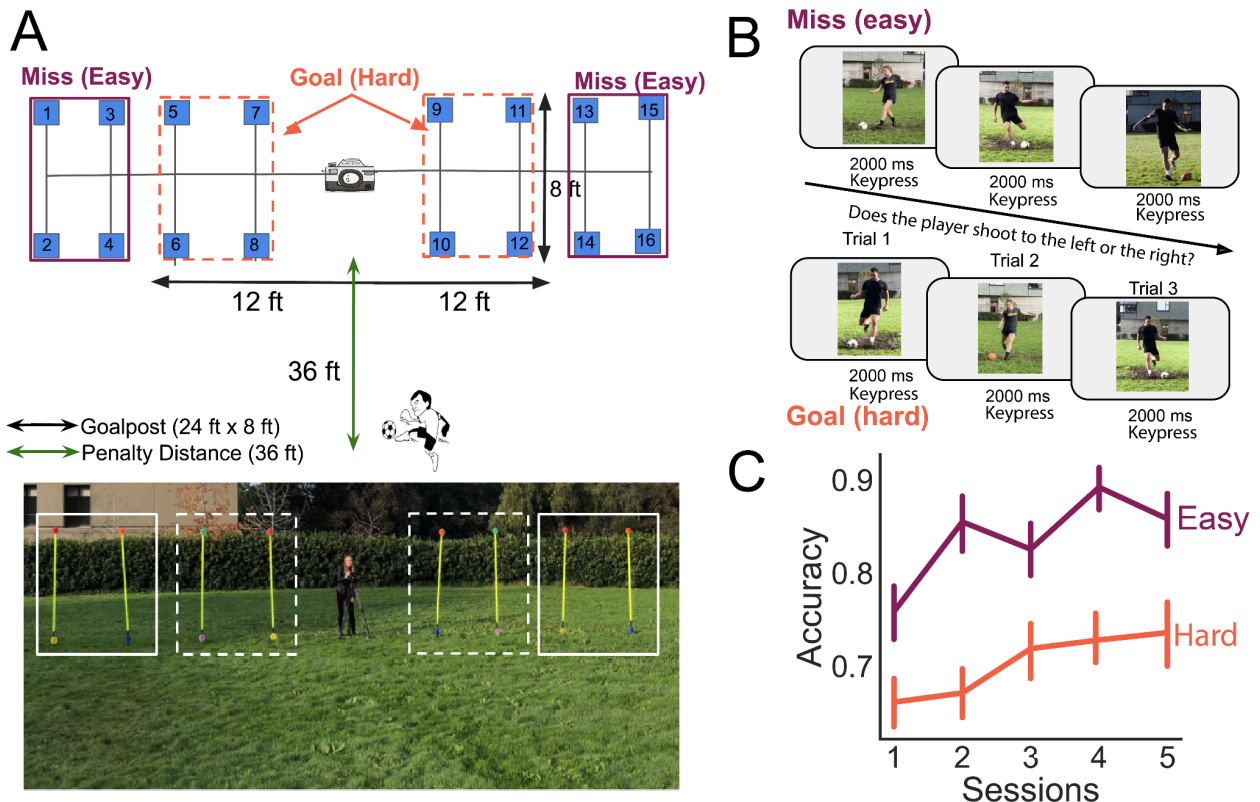
**Supplementary Figure 2. Task design for semantic prediction.** A) Two alternative forced choice task (2AFC), in which participants indicate if the target word comprises a semantically meaningful sentence. Words in the sentence base are presented sequentially for 500 ms after which time a fixation cross appears for 500 ms. The target word (“apple” or “class” in these example sentences) appears for 2000 ms and the participant indicates whether the sentence is either “meaningful” or “meaningless”. Sentence predictability was used to modulate difficulty, low cloze sentences are not easily predicted and high cloze sentences are strongly predicted. B) Reaction time performance across sessions for both easy and hard conditions.



**Supplementary Figure 3. Task design for N back.** A) Two alternative forced choice task (2AFC), in which participants indicate “True” if the current image on screen matches the image displayed two items previously (2-Back) or “False” if the current item does not match. Easy and hard conditions are indicated respectively by a current image that does and does not have a match two items previous. B) Reaction time performance across sessions for both easy and hard conditions.

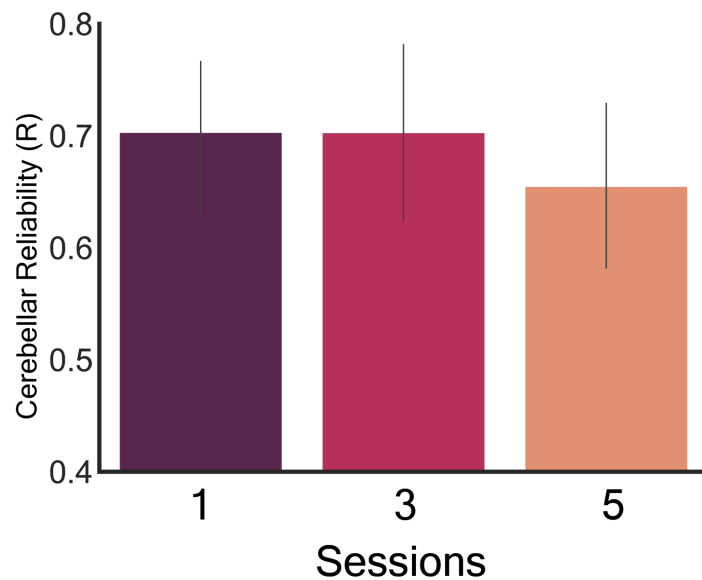


**Supplementary Figure 4. Task design for social prediction.** A) Two alternative forced choice task (2AFC), in which participants indicate “Hug” or “Handshake” if actors in a video will hug or shake hands at the end of the interaction. Task difficulty is modulated by the amount of time preceding the onset of the interaction. In the easy and hard conditions respectively, the video stops 200 ms and 500 ms before the actors interact (handshake: palms touch; hug: upper torsos touch). B) Accuracy performance across sessions for both easy and hard conditions.

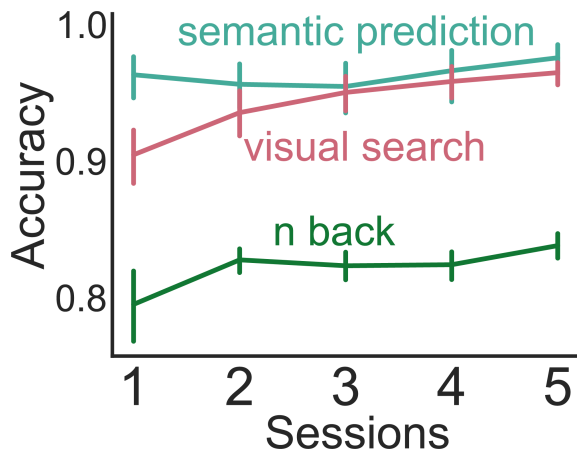


**Supplementary Figure 5. Task design for action prediction.** A) Construction of task stimuli: upper panel depicts a schematic of the layout of the 16 targets at which varsity

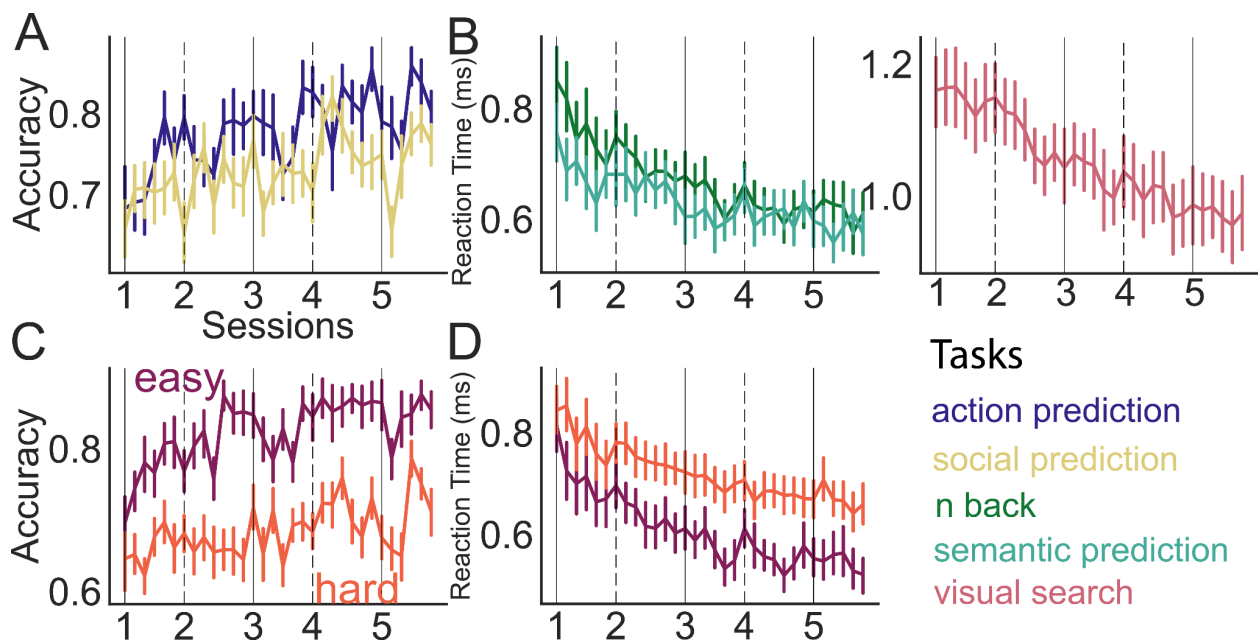
soccer players were instructed to kick the ball, and lower panel shows a photograph of the actual layout of the targets. The ball was positioned on a spot 36 ft from the goal line. On each trial, players were recorded as they ran up to the ball and took aim at one of the 16 instructed targets. Targets were constructed using 8 identical poles (8 ft in height), and two targets were placed on the upper and lower parts of each pole. Poles were spaced 3 ft apart from each other and from the video camera. Four poles (dashed lines) were used to represent “goals” (goal post is 24 ft in width x 8 ft in height), and the other four poles (solid lines) were used to represent “misses”. B) Two alternative forced choice task (2AFC), in which participants indicate “Left” or “Right” if they predict that a soccer player is going to kick a ball to the left or the right of the goalkeeper. Task difficulty is modulated by “Misses” and “Goals”. For example, if a soccer player aims for targets 5 thru 12, these shots qualify as goals and they are harder to predict, and if the player aims for targets 1 thru 4, and 13 thru 16, these qualify as misses, and they are easier to predict. B) Accuracy performance across sessions for both easy and hard conditions.



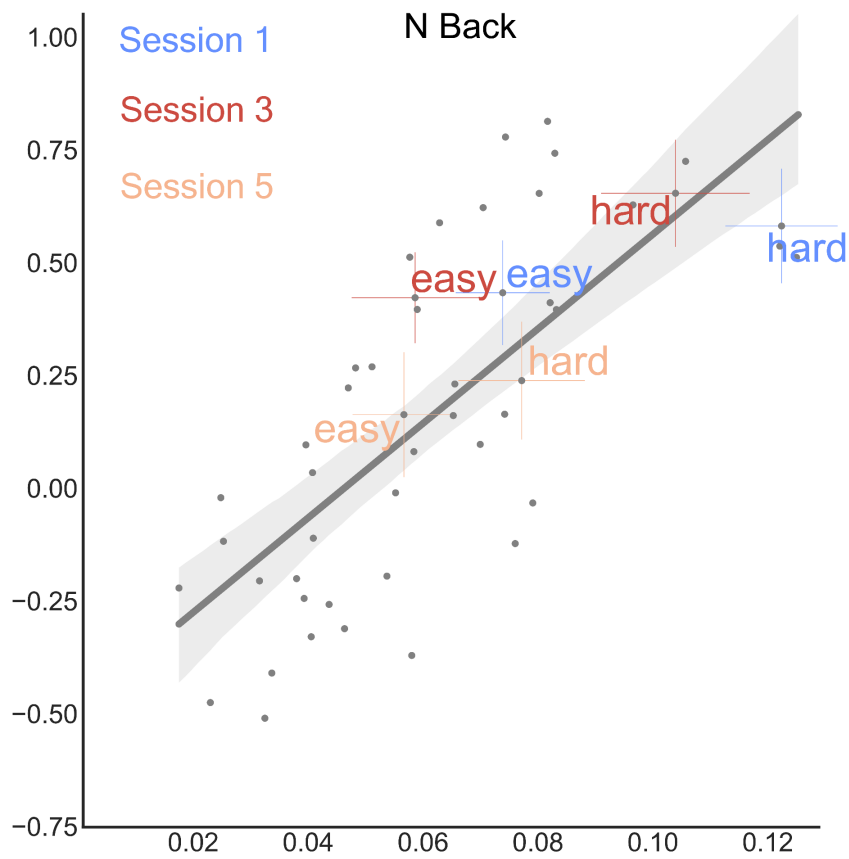
**Supplementary Figure 6. Reliability of cerebellar data.** A) Reliability of cerebellar activity patterns (Pearson’s correlation) for each scanning session (averaged across voxels). Reliability was calculated by correlating activity patterns of odd and even runs per session. B) Voxelwise reliability, averaged across all three sessions.



**Supplementary Figure 7. Accuracy across sessions for Visual Search, Semantic Prediction, and N Back tasks.** Behavioral performance for these three tasks were measured using reaction time. Performance on all tasks was on average >80% (N Back) and >90% (Visual Search, Semantic Prediction) for each session.



**Supplementary Figure 8. Performance across runs for all five tasks and levels of difficulty.** Accuracy on Action Prediction and Social Prediction (A) tasks improved across sessions, and reaction time on N Back, Semantic Prediction, and Visual Search (B) tasks decreased across sessions. Accuracy was greater for easy compared to hard conditions (averaged across Action Prediction and Social Prediction tasks), and reaction time was slower for hard compared to easy conditions (averaged across N Back, Semantic Prediction, and Visual Search tasks).



**Supplementary Figure 9. Cortico-cerebellar model predictions for the N Back task, across learning.** *The model underpredicts easy task conditions for all three sessions, and overpredicts hard conditions (with the exception of Session 3).*

# General Discussion

In this work, I set out to address fundamental questions about cortico-cerebellar contributions to cognition. My approach deviated from traditional methods, namely, I took advantage of tools and statistical techniques from the “Big Data” revolution to address important and open-ended questions about the cerebellum. I addressed three main topics in my work: the first was to capture how the cerebellum is functionally organized in humans, the second was to understand how the cerebral cortex is functionally connected to discrete regions of the cerebellum, and the final question set out to explore mechanistic and functional roles of the cerebellum in cognition.

I used a number of approaches to address each of these aims. To develop and evaluate a map of cerebellar function, I leveraged a multi-task battery that engaged a diverse set of cognitive processes (King et al. 2019). Then, to assess the functional relationship between the cerebral cortex and the cerebellum, I designed models of cortico-cerebellar connectivity using machine learning principles to assess convergence from the cortex to the cerebellum (King et al. 2022). Finally, to evaluate testable models of cerebellar function, I used the connectivity models I had previously established to predict changes in cerebellar activity patterns across different stages of learning. I also complemented the neuroimaging approach with a behavioral paradigm that tested two functional theories of the cerebellum in a patient population (King et al. 2022).

Overall, I believe the work I have presented offers key insights that will serve the neuroscientific community in doing both basic science and translational research. In Chapter 1, I presented a novel functional parcellation of the human cerebellum that summarizes cerebellar activation in response to a large number of domains from language and working memory to social cognition and action observation. The development of a functional atlas for the human cerebellum has already proven to be a useful addition for the community and has been adopted by several researchers (Walters et al. 2022; van Es et al. 2019). For example, researchers can use this parcellation to create meaningful region-of-interests (ROI) beyond the lobular compartmentation that currently exists. The second innovation is a metric I developed that can be used to evaluate the boundaries of any functional (or random) parcellation of the human brain. This metric has already been used, and extended in scope, to evaluate parcellations of the cerebral cortex (Zhi et al. 2022). Moreover, the overarching insight that arises from this work is a comprehensive understanding of how cognitive tasks, spanning many different domains, activate distinct functional regions of the cerebellar cortex. Importantly, this parcellation clearly demonstrates that a majority of the cerebellar cortex is activated by complex cognitive processes, while a smaller, yet distinct set of regions are activated by sensorimotor function. This observation serves as a strong corroboration of findings previously reported in the neuroimaging literature: the cerebellum is activated in response to many cognitive processes, unexplained by motor function. This insight is possible in large part to a feature-based encoding model that allows for the activation in each functional region to be interpreted in the context of task labels (e.g., cognitive features). Of course, there is no such thing as “one true parcellation” of the cerebellum, or any other brain region. Therefore, I hope that this

work is expanded and improved until a better and even more comprehensive parcellation of the human cerebellum is created.

In Chapter 2, I built a model with the goal of capturing the functional relationship between the cerebral cortex and the cerebellum. This approach is essential for testing theories about a cerebellar role across functional domains. Specifically, I compared how well cerebellar activity patterns in individual participants could be predicted from cortical activity patterns alone. I found that the most robust model of cortico-cerebellar connectivity allowed for a modest degree of convergence from cortical regions onto cerebellar voxels. Specifically, hand motor areas in lobules V/VI were well predicted by activity in a circumscribed region of the contralateral sensorimotor cortex. In contrast, the prediction of activity patterns in lobule VII (Crus I/II), a cerebellar area engaged during cognitive tasks, required inputs from prefrontal, parietal and temporal regions. These findings point to important differences in the way that functional subdivisions of the cerebellar cortex work in concert with the neocortex to support motor and cognitive function.

In Chapters 3 and 4, I evaluated prominent functional and mechanistic theories of the cerebellum's involvement in cognition using neuropsychology and neuroimaging. In Chapter 3, I tested a popular theory of cerebellar function, the internal model hypothesis, to determine whether individuals with cerebellar degeneration had selective deficits in semantic prediction. Using a sentence verification task, I compared reaction time performance across sentences that were strongly or weakly predictable in two groups: patients with cerebellar degeneration, and healthy controls. I 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 find support for either hypothesis, challenging current theorizing about the role of the cerebellum in language processing, and pointing to a misalignment between neuroimaging and neuropsychology research on this topic.

In Chapter 4, I leveraged the detailed cortico-cerebellar connectivity model established in Chapter 2 to determine whether the cerebellum uniquely contributes to cognitive function across learning. To determine whether there were task-specific deviations from the linear connectivity model, I compared observed and predicted cerebellar activity patterns for each of 5 unique tasks. I found tasks that engaged dynamic action predictions were underpredicted by the connectivity model, while perceptual and working memory tasks were overpredicted, suggesting a unique functional role for the cerebellum in action prediction. Importantly, while the cerebellar BOLD signal decreased across learning for all domains, the only tasks that were underpredicted by the cerebellum at each stage of learning were processes that engaged dynamic action prediction. These findings suggest that the cerebellum may play a task-specific role in coordinating cortico-cortico interactions during early stages of learning.

It is my hope that the results from this work will be useful to other neuroscience researchers who seek to understand the fundamental contributions that the cerebellum makes to human cognition. Furthermore, I hope that in the (not so distant) future, some of the results and insights I've presented in this work will be leveraged to develop

diagnoses, treatments, and interventions for the physiological and behavioral symptoms that arise from cerebellar disease, pathology, and developmental abnormalities.

## 6.1 References

### 6.1.1 Chapter 1

1. Ivry RB. & Baldo JV Is the cerebellum involved in learning and cognition? *Curr. Opin. Neurobiol.* 2, 212–216 (1992).
2. Kelly RM & Strick PL Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci.* 23, 8432–8444 (2003).
3. Allen G, Buxton RB, Wong EC & Courchesne E. Attentional activation of the cerebellum independent of motor involvement. *Science* 275, 1940–1943 (1997).
1. Larsell O. The development of the cerebellum in man in relation to its comparative anatomy. *J. Comp. Neurol.* 87, 85–129 (1947).
2. Buckner RL, Krienen FM, Castellanos A, Diaz JC & Yeo BTT The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 2322–2345 (2011).
3. Ji JL et al. Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* 185, 35–57 (2019).
4. Guell X, Schmahmann JD, Gabrieli J. & Ghosh SS Functional gradients of the cerebellum. *eLife* 7, e36652 (2018).
5. Stoodley CJ, Valera EM & Schmahmann JD Functional topography of the cerebellum for motor and cognitive tasks: an fMRI study. *Neuroimage* 59, 1560–1570 (2012).
6. Stoodley CJ & Schmahmann JD Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage* 44, 489–501 (2009).
7. Wiestler T, McGonigle DJ & Diedrichsen J. Integration of sensory and motor representations of single fingers in the human cerebellum. *J. Neurophysiol.* 105, 3042–3053 (2011).
8. Ohtsuka K. & Noda H. Discharge properties of Purkinje cells in the oculomotor vermis during visually guided saccades in the macaque monkey. *J. Neurophysiol.* 74, 1828–1840 (1995).
9. Nitschke MF et al. Activation of cerebellar hemispheres in spatial memorization of saccadic eye movements: an fMRI study. *Hum. Brain Mapp.* 22, 155–164 (2004).
10. Horak FB & Diener HC Cerebellar control of postural scaling and central set in stance. *J. Neurophysiol.* 72, 479–493 (1994).
11. Diedrichsen J. & Zotow E. Surface-based display of volume-averaged cerebellar imaging data. *PLoS ONE* 10, e0133402 (2015).
12. Nguyen VT et al. Distinct cerebellar contributions to cognitive- perceptual dynamics during natural viewing. *Cereb. Cortex* 27, 5652–5662 (2017).
13. Barch DM et al. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–189 (2013).

14. Diedrichsen J. & Kriegeskorte N. Representational models: a common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Comput. Biol.* 13, e1005508 (2017).
15. Poldrack RA et al. The Cognitive Atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform.* 5, 17 (2011).
16. Marek S. et al. Spatial and temporal organization of the individual human cerebellum. *Neuron* 100, 977–993.e7 (2018).
17. Diedrichsen J, Balsters JH, Flavell J, Cussans E. & Ramnani N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* 46, 39–46 (2009).
18. Airey DC, Lu L. & Williams RW Genetic control of the mouse cerebellum: identification of quantitative trait loci modulating size and architecture. *J. Neurosci.* 21, 5099–5109 (2001).
19. Apps R. & Hawkes R. Cerebellar cortical organization: a one-map hypothesis. *Nat. Rev. Neurosci.* 10, 670–681 (2009).
20. Sugihara I. & Shinoda Y. Molecular, topographic, and functional organization of the cerebellar cortex: a study with combined aldolase C and olivocerebellar labeling. *J. Neurosci.* 24, 8771–8785 (2004).
21. Leclerc N, Doré L, Parent A. & Hawkes R. The compartmentalization of the monkey and rat cerebellar cortex: zebrin I and cytochrome oxidase. *Brain Res.* 506, 70–78 (1990).
22. Lauritzen M. Relationship of spikes, synaptic activity, and local changes of cerebral blood flow. *J. Cereb. Blood Flow Metab.* 21, 1367–1383 (2001).
23. Hawkes R, Gallagher E. & Ozol K. Blebs in the mouse cerebellar granular layer as a sign of structural inhomogeneity. 1. Anterior lobe vermis. *Acta Anat. (Basel)* 158, 205–214 (1997).
24. Tavor I. et al. Task-free MRI predicts individual differences in brain activity during task performance. *Science* 352, 216–220 (2016).
25. Martin TA, Keating JG, Goodkin HP, Bastian AJ & Thach WT Throwing while looking through prisms. I. Focal olivocerebellar lesions impair adaptation. *Brain* 119, 1183–1198 (1996).
26. Andreasen NC & Pierson R. The role of the cerebellum in schizophrenia. *Biol. Psychiatry* 64, 81–88 (2008).
27. Moberget T. et al. Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Mol. Psychiatry* 23, 1512–1520 (2018).
28. Cornelissen FW & Peters EM The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behav. Res. Methods Instrum. Comput.* 34, 613–617 (2002).
29. Ashburner J. et al. SPM12 Manual  
[www.fil.ion.ucl.ac.uk/spm/doc/spm12\\_manual.pdf](http://www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf) (2019).
30. Van Essen DC Cortical cartography and Caret software. *Neuroimage* 62, 757–764 (2012).
31. Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. *Neuroimage* 33, 127–138 (2006).
32. Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113 (2007).

33. Sereno MI, Diedrichsen J, Tachrouf M, Silva G. & De Zeeuw C. Reconstruction and unfolding of the human cerebellar cortex from high-resolution post-mortem MRI. Society for Neuroscience, abstr. 733 (2014).
34. Oosterhof NN, Wiestler T, Downing PE & Diedrichsen J. A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage* 56, 593–600 (2011).
35. Ding CH, Li T. & Jordan MI Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 45–55 (2010).
36. Walther A. et al. Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* 137, 188–200 (2016).
37. Boly, M. *et al.* When thoughts become action : An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients. *Hum. Brain Mapp. J.* **36**, 979–992 (2007). Egner, T. & Hirsch, J. The neural correlates and functional integration of cognitive control in a Stroop task. **24**, 539–547 (2005).
38. Owen, A. M., McMillan, K. M., Laird, A. R. & Bullmore, E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* **25**, 46–59 (2005).
39. Schubotz, R. I., & von Cramon, D. Y. Interval and ordinal properties of sequences are associated with distinct premotor areas. *Cerebral Cortex* **11**, 210-222 (2001).
40. Rickard, T. C., Romero, S. G., Basso, G., Wharton, C., Flitman, S., & Grafman, J. The calculating brain: an fMRI study. *Neuropsychologia* **38**, 325-335 (2000).
41. Moulton, E. A., Elman, I., Pendse, G., Schmahmann, J., Becerra, L., & Borsook, D. Aversion- related circuitry in the cerebellum: responses to noxious heat and unpleasant images. *Journal of neuroscience* **31**, 3795-3804 (2011).
42. Elliott, R., Rubinsztein, J. S., Sahakian, B. J., & Dolan, R. J. Selective attention to emotional stimuli in a verbal go/no-go task: an fMRI study. *Neuroreport* **11** 1739-1744 (2000). Dodell-feder, D., Koster-hale, J., Bedny, M. & Saxe, R. NeuroImage fMRI item analysis in a theory of mind task. *Neuroimage* **55**, 705–712 (2011).
43. Fiez, J. A. Cerebellar Contributions to Cognition. *Neuron* **16**, 13–15 (1996).
44. Wiestler, T. & Diedrichsen, J. Skill learning strengthens cortical representations of motor sequences. *Elife* **2013**, 1–20 (2013).
45. Cross, E. S., Cohen, N. R., Hamilton, A. F. D. C., Ramsey, R., Wolford, G., & Grafton, S. T. Physical experience leads to enhanced object perception in parietal cortex: insights from knot tying. *Neuropsychologia* **50**, 3207-3217 (2012).
46. Donner, T. H., Kettermann, A., Diesch, E., Ostendorf, F., Villringer, A., & Brandt, S. A. Visual feature and conjunction searches of equal difficulty engage only partially overlapping frontoparietal networks. *Neuroimage* **15**, 16-25 (2002).
47. Shepard, R. N. & Metzler, J. Mental Rotation of Three-Dimensional Objects. *Science* (80-. ). **171**, 701–703 (1971).
48. Ganis, G. & Kievit, R. A. A new set of three-dimensional shapes for investigating mental rotation processes: validation data and stimulus set. *J. Chem. Inf. Model.* **53**, 1689–1699 (2013).
49. Troje, N. F. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision* **2**, 2-2 (2002).

50. Ito, T., Kulkarni, K. R., Schultz, D. H., Mill, R. D., Chen, R. H., Solomyak, L. I., & Cole, M. W. Cognitive task information is transferred between brain regions via resting-state network topology. *Nature communications* **8**, 1027 (2017).
51. Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B. & Endestad, T. Generalized Role for the Cerebellum in Encoding Internal Models: Evidence from Semantic Processing. *J. Neurosci.* **34**, 2871–2878 (2014).
52. D'Mello, A. M., Turkeltaub, P. E., & Stoodley, C. J. Cerebellar tDCS modulates neural circuits during semantic prediction: a combined tDCS-fMRI study. *Journal of Neuroscience* **37**, 1604– 1613 (2017).
53. Bischoff-Grethe, A., Ivry, R. B., & Grafton, S. T. Cerebellar involvement in response reassignment rather than attention. *Journal of Neuroscience* **22**, 546–553 (2002).
54. Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B. & Endestad, T. Generalized Role for the Cerebellum in Encoding Internal Models: Evidence from Semantic Processing. *J. Neurosci.* **34**, 2871–2878 (2014).
55. D'Mello, A. M., Turkeltaub, P. E., & Stoodley, C. J. Cerebellar tDCS modulates neural circuits during semantic prediction: a combined tDCS-fMRI study. *Journal of Neuroscience* **37**, 1604– 1613 (2017).
56. Bischoff-Grethe, A., Ivry, R. B., & Grafton, S. T. Cerebellar involvement in response reassignment rather than attention. *Journal of Neuroscience* **22**, 546–553 (2002).

## 6.1.2 Chapter 2

1. Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. Positron emission tomographic studies of the processing of single words. *J. Cogn. Neurosci.* **1**, 153–170 (1989).
2. Allen, G., Buxton, R. B., Wong, E. C. & Courchesne, E. Attentional activation of the cerebellum independent of motor involvement. *Science* **275**, 1940–1943 (1997).
3. Bischoff-Grethe, A., Ivry, R. B. & Grafton, S. T. Cerebellar involvement in response reassignment rather than attention. *J. Neurosci.* **22**, 546–553 (2002).
4. Balsters, J. H., Whelan, C. D., Robertson, I. H. & Ramnani, N. Cerebellum and cognition: evidence for the encoding of higher order rules. *Cereb. Cortex* **23**, 1433–1443 (2013).
5. Marvel, C. L. & Desmond, J. E. Functional topography of the cerebellum in verbal working memory. *Neuropsychol. Rev.* **20**, 271–279 (2010).
6. Fiez, J. A. The cerebellum and language: Persistent themes and findings. *Brain Lang.* **161**, 1–3 (2016).
7. Van Overwalle, F., D'aes, T. & Mariën, P. Social cognition and the cerebellum: A meta-analytic connectivity analysis. *Hum. Brain Mapp.* **36**, 5137–5154 (2015).
8. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat. Neurosci.* **22**, 1371–1378 (2019).

9. Kelly, R. M. & Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci.* **23**, 8432–8444 (2003).
10. Dum, R. P. & Strick, P. L. An Unfolded Map of the Cerebellar Dentate Nucleus and its Projections to the Cerebral Cortex. *J. Neurophysiol.* **89**, 634–639 (2003).
11. Strick, P. L., Dum, R. P. & Fiez, J. A. Cerebellum and nonmotor function. *Annu. Rev. Neurosci.* **32**, 413–434 (2009).
12. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).
13. Ji, J. L. *et al.* Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* **185**, 35–57 (2019).
14. Marek, S. *et al.* Spatial and Temporal Organization of the Individual Human Cerebellum. *Neuron* **100**, 977–993.e7 (2018).
15. Henschke, J. U. & Pakan, J. M. P. Disynaptic cerebrocerebellar pathways originating from multiple functionally distinct cortical areas. *Elife* **9**, e59148 (2020).
16. Pisano, T. J. *et al.* Homologous organization of cerebellar pathways to sensory, motor, and associative forebrain. *Cell Rep.* **36**, 109721 (2021).
17. Huang, C.-C. *et al.* Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. *Elife* **2**, e00400 (2013).
18. Yeo, B. T. T. *et al.* Functional Specialization and Flexibility in Human Association Cortex. *Cereb. Cortex* **25**, 3654–3672 (2015).
19. Yeo, B. T. T., Krienen, F. M., Chee, M. W. L. & Buckner, R. L. Estimates of segregation and overlap of functional connectivity networks in the human cerebral cortex. *Neuroimage* **88**, 212–227 (2014).
20. Bertolero, M. A., Yeo, B. T. T. & D'Esposito, M. The modular and integrative functional architecture of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6798–807 (2015).
21. Cole, M. W., Ito, T., Bassett, D. S. & Schultz, D. H. Activity flow over resting-state networks shapes cognitive task activations. *Nat. Neurosci.* **19**, 1718–1726 (2016).
22. Mell, M. M., St-Yves, G. & Naselaris, T. Voxel-to-voxel predictive models reveal unexpected structure in unexplained variance. *Neuroimage* **238**, 118266 (2021).
23. Mathiesen, C., Caesar, K. & Lauritzen, M. Temporal coupling between neuronal activity and blood flow in rat cerebellar cortex as indicated by field potential analysis. *J. Physiol.* **523 Pt 1**, 235–246 (2000).
24. Thomsen, K., Offenhauser, N. & Lauritzen, M. Principal neuron spiking: neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *J. Physiol.* **560**, 181–189 (2004).
25. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
26. Zhi, D., King, M., Hernandez-Castillo, C. R. & Diedrichsen, J. Evaluating brain parcellations using the distance-controlled boundary coefficient. *Hum. Brain Mapp.* (2022) doi:10.1002/hbm.25878.
27. Krienen, F. M. & Buckner, R. L. Segregated fronto-cerebellar circuits revealed by intrinsic functional connectivity. *Cereb. Cortex* **19**, 2485–2497 (2009).

28. O'Reilly, J. X., Beckmann, C. F., Tomassini, V., Ramnani, N. & Johansen-Berg, H. Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity. *Cereb. Cortex* **20**, 953–965 (2010).
29. Guell, X., Schmahmann, J. D., Gabrieli, J. D. E. & Ghosh, S. S. Functional gradients of the cerebellum. *Elife* **7**, e36652 (2018).
30. Balsters, J. H. *et al.* Evolution of the cerebellar cortex: the selective expansion of prefrontal-projecting cerebellar lobules. *Neuroimage* **49**, 2045–2052 (2010).
31. Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M. & Ivry, R. B. Universal Transform or Multiple Functionality? Understanding the Contribution of the Human Cerebellum across Task Domains. *Neuron* **102**, 918–928 (2019).
32. De Zeeuw, C. I., Lisberger, S. G. & Raymond, J. L. Diversity and dynamism in the cerebellum. *Nat. Neurosci.* **24**, 160–167 (2021).
33. Tavor, I. *et al.* Task-free MRI predicts individual differences in brain activity during task performance. *Science* **352**, 216–220 (2016).
34. Thomsen, K., Piilgaard, H., Gjedde, A., Bonvento, G. & Lauritzen, M. Principal cell spiking, postsynaptic excitation, and oxygen consumption in the rat cerebellar cortex. *J. Neurophysiol.* **102**, 1503–1512 (2009).
35. Schwarz, C. & Thier, P. Binding of signals relevant for action: towards a hypothesis of the functional role of the pontine nuclei. *Trends Neurosci.* **22**, 443–451 (1999).
36. Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S. & Petersen, S. E. Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**, 238–251 (2014).
37. Braun, U. *et al.* Dynamic reconfiguration of frontal brain networks during executive cognition in humans. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11678–11683 (2015).
38. Allen, E. A. *et al.* Tracking Whole-Brain Connectivity Dynamics in the Resting State. *Cereb. Cortex* **24**, 663–676 (2012).
39. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
40. Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J. & Coalson, T. Parcellations and Hemispheric Asymmetries of Human Cerebral Cortex Analyzed on Surface-Based Atlases. *Cereb. Cortex* **22**, 2241–2262 (2011).
41. Diedrichsen, J. A spatially unbiased atlas template of the human cerebellum. *Neuroimage* **33**, 127–138 (2006).
42. Diedrichsen, J. & Zotow, E. Surface-Based Display of Volume-Averaged Cerebellar Imaging Data. *PLoS One* **10**, e0133402 (2015).
43. Mumford, J. A., Turner, B. O., Ashby, F. G. & Poldrack, R. A. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* **59**, 2636–2643 (2012).
44. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).
45. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
46. Esteban, O. *et al.* Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nat. Protoc.* **15**, 2186–2202 (2020).

47. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2017).
48. Arslan, S., Parisot, S. & Rueckert, D. Joint Spectral Decomposition for the Parcellation of the Human Cerebral Cortex Using Resting-State fMRI. in *Information Processing in Medical Imaging* 85–97 (Springer International Publishing, 2015). doi:10.1007/978-3-319-19992-4\_7.
49. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
50. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex* **26**, 288–303 (2016).
51. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* **82**, 403–415 (2013).
52. Zhi, D., King, M., Hernandez-Castillo, C. R. & Diedrichsen, J. Evaluating brain parcellations using the distance controlled boundary coefficient. *bioRxiv* 2021.05.11.443151 (2021) doi:10.1101/2021.05.11.443151.

### 6.1.3 Chapter 3

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.
17. 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.
18. 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.
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. Ivry, R. B., & Fiez, J. A. (2000). Cerebellar contributions to cognition and imagery. *The new cognitive neurosciences*, 2, 999-1011.

29. 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.
30. Leiner, H. C., Leiner, A. L. & Dow, R. S. (1986). Does the Cerebellum Contribute to Mental Skills? *Behavioral Neuroscience*, 100(44), 443–454.
31. 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.
32. Lesage, E., Hansen, P. C., & Miall, R. C. (2017). Right lateral cerebellum represents linguistic predictability. *Journal of Neuroscience*, 37(26), 6231-6241.
33. 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.
34. 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.
35. 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.
36. Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.
37. Marvel, C. L., & Desmond, J. E. (2010). Functional topography of the cerebellum in verbal working memory. *Neuropsychology review*, 20(3), 271-279.
38. 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).
39. 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.
40. 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.
41. 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.
42. 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.
43. Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, 116(2), 220.
44. 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*.

45. 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.
46. Pickering, M. J. & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110.
47. Poort, E. D., & Rodd, J. M. (2019). Towards a distributed connectionist account of cognates and interlingual homographs: Evidence from semantic relatedness tasks. *PeerJ*, 7.
48. 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.
49. 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.
50. 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.
51. Riva, D. (1998). Language deficits in a child with unilateral (left) temporo-basal and cerebellar lesions. *Neuropsychologia*, 36(1), 71-75.
52. Schmahmann, J. D. (1996). From movement to thought: anatomic substrates of the cerebellar contribution to cognitive processing. *Human Brain Mapping*, 4(3), 174-198.
53. Schmahmann, J. D. (1998). Dysmetria of thought: clinical consequences of cerebellar dysfunction on cognition and affect. *Trends in Cognitive Sciences*, 2(9), 362-371.
54. Schmahmann, J. D. & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain*, 121(4), 561–579.
55. 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.
56. Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701-703.
57. Smith NJ, Levy R. Cognition. 2013 Sep;128(3):302-19.
58. 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.
59. 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.
60. 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.

61. Stoodley, C. J., & Schmahmann, J. D. (2009). The cerebellum and language: evidence from patients with cerebellar degeneration. *Brain and language*, 110(3), 149-153.
62. 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.
63. 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.
64. Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880-1882.
65. Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9), 338-347.

## 6.1.4 Chapter 4

1. Ackermann, H., Gräber, S., Hertrich, I., & Daum, I. (1997). Categorical speech perception in cerebellar disorders. *Brain and language*, 60(2), 323-331.
2. Allen, G., Buxton, R. B., Wong, E. C., & Courchesne, E. (1997). Attentional activation of the cerebellum independent of motor involvement. *Science*, 275(5308), 1940-1943.
3. Allen GI, Tsukahara N. Cerebrocerebellar communication systems. *Physiol Rev*. 1974;54:957–1006.
4. Balsters, J. H., & Ramnani, N. (2011). Cerebellar plasticity and the automation of first-order rules. *Journal of Neuroscience*, 31(6), 2305-2312.
5. Balsters, J. H. & Ramnani, N. Cerebellar Plasticity and the Automation of First-Order Rules. **31**, 2305–2312 (2011).
6. Barton, R. A. (2012). Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1599), 2097-2107.
7. Blakemore, S. J., & Sirigu, A. (2003). Action prediction in the cerebellum and in the parietal lobe. *Experimental Brain Research*, 153(2), 239-245.
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. Bischoff-Grethe, A., Ivry, R. B., & Grafton, S. T. (2002). Cerebellar involvement in response reassignment rather than attention. *Journal of Neuroscience*, 22(2), 546-553.
11. Chan, H. Y., Smidts, A., Schoots, V. C., Dietvorst, R. C., & Boksem, M. A. (2019). Neural similarity at temporal lobe and cerebellum predicts out-of-sample preference and recall for video stimuli. *Neuroimage*, 197, 391-401.
12. Desmond, J. E., Gabrieli, J. D., Wagner, A. D., Ginier, B. L. & Glover, G. H. Lobular patterns of cerebellar activation in verbal working-memory and

- finger-tapping tasks as revealed by functional MRI. *J. Neurosci.* 17, 9675–9685 (1997).
13. De Smet, H. J., Paquier, P., Verhoeven, J., & Mariën, P. (2013). The cerebellum: its role in language and related cognitive and affective functions. *Brain and language*, 127(3), 334–342.
  14. Dum, R. P. & Strick, P. L. An Unfolded Map of the Cerebellar Dentate Nucleus and its Projections to the Cerebral Cortex. *J. Neurophysiol.* **89**, 634–639 (2003).
  15. D'Mello, A. M., Gabrieli, J. D., & Nee, D. E. (2020). Evidence for hierarchical cognitive control in the human cerebellum. *Current Biology*, 30(10), 1881–1892.
  16. Errante, A., & Fogassi, L. (2020). Activation of cerebellum and basal ganglia during the observation and execution of manipulative actions. *Scientific reports*, 10(1), 1–15.
  17. Grafton, S. T. Woods, R. P. & Tyszka, M. (1993). Functional imaging of procedural motor learning: relating cerebral blood flow with individual subject performance. *Hum. Brain Mapp.* 1, 221–234.
  18. Guell, X., Schmahmann, J. D., Gabrieli, J. D., & Ghosh, S. S. (2018). Functional gradients of the cerebellum. *Elife*, 7, e36652.
  19. Guell, X., Gabrieli, J. D., & Schmahmann, J. D. (2018). Embodied cognition and the cerebellum: perspectives from the dysmetria of thought and the universal cerebellar transform theories. *Cortex*, 100, 140–148.
  20. Heleven, E., Van Dun, K., De Witte, S., Baeken, C., & Van Overwalle, F. (2021). The role of the cerebellum in social and non-social action sequences: a preliminary LF-rTMS study. *Frontiers in Human Neuroscience*, 15, 85.
  21. Imamizu, H. et al. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature* 403, 192–195 (2000)
  22. Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, 9(4), 304–313.
  23. Ivry, Richard B. Keele, S. Timing Functions of The Cerebellum. *J. Cogn. Neurosci.* (1989).
  24. Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. & Passingham, R. E. Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *J. Neurophysiol.* 77, 1325–1337 (1997).
  25. Kelly, R. M. & Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci.* **23**, 8432–8444 (2003).
  26. 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.
  27. Leggio, M., & Molinari, M. (2015). Cerebellar sequencing: a trick for predicting the future. *The Cerebellum*, 14(1), 35–38.
  28. 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.
  29. Lesage, E., Hansen, P. C., & Miall, R. C. (2017). Right lateral cerebellum represents linguistic predictability. *Journal of Neuroscience*, 37(26), 6231–6241.
  30. Marvel, C. L., & Desmond, J. E. (2010). The contributions of cerebro-cerebellar circuitry to executive verbal working memory. *Cortex*, 46(7), 880–895.

31. 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).
32. Miall, R. C. & Jenkinson, E. W. Functional imaging of changes in cerebellar activity related to learning during a novel eye-hand tracking task. *Exp. Brain Res.* 166, 170–183 (2005).
33. 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. *Journal of Neuroscience*, 34(8), 2871-2878.
34. Middleton, F. A., & Strick, P. L. (2001). Cerebellar projections to the prefrontal cortex of the primate. *Journal of neuroscience*, 21(2), 700-712.
35. Nguyen, V. T., Sonkusare, S., Stadler, J., Hu, X., Breakspear, M., & Guo, C. C. (2017). Distinct cerebellar contributions to cognitive-perceptual dynamics during natural viewing. *Cerebral Cortex*, 27(12), 5652-5662.
36. O'Reilly, J. X., Beckmann, C. F., Tomassini, V., Ramnani, N., & Johansen-Berg, H. (2010). Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity. *Cerebral cortex*, 20(4), 953-965.
37. 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*.
38. Schmahmann, J. D., & Pandya, D. N. (1997). Anatomic organization of the basilar pontine projections from prefrontal cortices in rhesus monkey. *Journal of Neuroscience*, 17(1), 438-458.
39. Schmahmann, J. D. (2004). Disorders of the cerebellum: ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. *The Journal of neuropsychiatry and clinical neurosciences*, 16(3), 367-378.
40. Shulman, G. L., Corbetta, M., Buckner, R. L., Fiez, J. A., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: I. Increases in subcortical structures and cerebellum but not in nonvisual cortex. *Journal of cognitive neuroscience*, 9(5), 624-647.
41. Schlerf, J., Ivry, R. B., & Diedrichsen, J. (2012). Encoding of sensory prediction errors in the human cerebellum. *Journal of Neuroscience*, 32(14), 4913-4922.
42. Schwarz, C., & Thier, P. (1999). Binding of signals relevant for action: towards a hypothesis of the functional role of the pontine nuclei. *Trends in neurosciences*, 22(10), 443-451.
43. 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.
44. Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: Fast, enduring, but non-specific. *Vision research*, 35(14), 2037-2043.
45. 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
46. Stojanoski, B., Lyons, K. M., Pearce, A. A., & Owen, A. M. (2018). Targeted training: Converging evidence against the transferable benefits of online brain training on cognitive function. *Neuropsychologia*, 117, 541-550.

47. Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*, 44(2), 489-501.
48. 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.
49. Stoodley, C. J., & Tsai, P. T. (2021). Adaptive prediction for social contexts: the cerebellar contribution to typical and atypical social behaviors. *Annual Review of Neuroscience*, 44, 475-493.
50. Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. *Neuroimage*, 86, 554-572.
51. Van Overwalle, F., Manto, M., Leggio, M., & Delgado-García, J. M. (2019). The sequencing process generated by the cerebellum crucially contributes to social interactions. *Medical Hypotheses*, 128, 33-42.
52. Wang, S. S. H., Kloth, A. D., & Badura, A. (2014). The cerebellum, sensitive periods, and autism. *Neuron*, 83(3), 518-532.
53. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347 (1998).

## 6.1.5 General Introduction and Discussion

1. Albus, S. A theory of cerebellar function. **10**, 25–61 (1971).
2. 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.
3. Andreasen, N. C., & Pierson, R. (2008). The role of the cerebellum in schizophrenia. *Biological psychiatry*, 64(2), 81-88.
4. Antonova, E., Sharma, T., Morris, R., & Kumari, V. (2004). The relationship between brain structure and neurocognition in schizophrenia: a selective review. *Schizophrenia research*, 70(2-3), 117-145.
5. Apps, R. & Hawkes, R. (2009). Cerebellar cortical organization: A one-map hypothesis. *Nature Reviews Neuroscience*, 10(9), 670-681.
6. Attwell, D., Iadecola, C. & Attwell, D. The neural basis of functional brain imaging signals. **25**, 621–625 (2002).
7. Baier, B., Dieterich, M., Stoeter, P., Birklein, F., & Müller, N. G. (2010). Anatomical correlate of impaired covert visual attentional processes in patients with cerebellar lesions. *Journal of Neuroscience*, 30(10), 3770-3776.
8. 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.
9. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 2322–2345 (2011).
10. Casini, L. & Ivry, R. B. Effects of divided attention on temporal processing in patients with lesions of the cerebellum or frontal lobe. *Neuropsychology* (1999).

11. Criscimagna-hemminger, S. E., Bastian, A. J. & Shadmehr, R. Size of Error Affects Cerebellar Contributions to Motor Learning. 2275–2284 (2010).
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. & Zotow, E. Surface-based display of volume-averaged cerebellar imaging data. *PLoS One* **10**, 1–18 (2015).
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. Eccles J. C., Ito M., Szentágothai J. (1967). *The cerebellum as a neuronal machine*. Oxford, England: Springer-Verlag.
16. 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. Guell, X., Schmahmann, J. D., Gabrieli, J. D., & Ghosh, S. S. (2018). Functional gradients of the cerebellum. *Elife*, 7, e36652.
18. 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.
19. Henschke, J. U., & Pakan, J. M. (2020). Disynaptic cerebrocerebellar pathways originating from multiple functionally distinct cortical areas. *Elife*, 9, e59148.
20. Huang, C., Sugino, K., Shima, Y., Guo, C. & Bai, S. Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. 1–17 (2013). doi:10.7554/eLife.00400.
21. Ito, M. (1993). Movement and thought: identical control mechanisms by the cerebellum.
22. Ivry, R. B., Keele, S. W., & Diener, H. C. (1988). Dissociation of the lateral and medial cerebellum in movement timing and movement execution. *Experimental brain research*, 73(1), 167-180.
23. Ivry, Richard B. Keele, S. Timing Functions of The Cerebellum. *J. Cogn. Neurosci.* (1989).
24. Ivry\*, R. B. & Diener, H. C. Impaired Velocity Perception in Patients with Lesions of the Cerebellum. *J. Cogn. Neurosci.* **3**, 355–366 (1991).
25. 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.
26. Kelly, R. M. & Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci.* **23**, 8432–8444 (2003).
27. 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.
28. King, M., Shahshahani, L., Ivry, R., & Diedrichsen, J. (2022). A general model of cortico-cerebellar connectivity, BioArxiv (2022).

29. King, M., Bruinsma, S., Ivry, R. (2022). No evidence for semantic prediction deficits in patients with spinocerebellar ataxia, *Neurobiology of Language* (submitted)(2022).
30. Leiner, H. C., Leiner, A. L. & Dow, R. S. Does the cerebellum contribute to mental skills? *Behav. Neurosci.* **100**, 443–454 (1986).
31. Larsell, O. The development of the cerebellum in man in relation to its comparative anatomy. *J. Comp. Neurol.* (1947).
32. 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.
33. Lesage, E., Hansen, P. C., & Miall, R. C. (2017). Right lateral cerebellum represents linguistic predictability. *Journal of Neuroscience*, **37**(26), 6231–6241.
34. Llinás, R. R. (1975). The cortex of the cerebellum. *Scientific American*, **232**(1), 56–71.
35. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. (2001).
36. Marr, B. Y. D. & Thach, T. A theory of cerebellar cortex. 437–470 (1969).
37. 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).
38. 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.
39. Moberget, T. & Ivry, R. B. Cerebellar contributions to motor control and language comprehension: Searching for common computational principles. *Ann. N. Y. Acad. Sci.* **1369**, 154–171 (2016).
40. O'Reilly, J. X., Beckmann, C. F., Tomassini, V., Ramnani, N., & Johansen-Berg, H. (2010). Distinct and overlapping functional zones in the cerebellum defined by resting state functional connectivity. *Cerebral cortex*, **20**(4), 953–965.
41. Peterburs, J., Blevins, L. C., Sheu, Y. S., & Desmond, J. E. (2019). Cerebellar contributions to sequence prediction in verbal working memory. *Brain Structure and Function*, **224**(1), 485–499.
42. Petersen, S. E., Fox, P. T., Posner, M., Mintun, M. & Raichle, M. Positron emission tomographic studies of the processing of single words. *J. Cogn. Neurosci.* (1989). doi:10.1162/jocn.1989.1.2.153.
43. Riva, D. (1998). Language deficits in a child with homolateral (left) temporo-basal and cerebellar lesions. *Neuropsychologia*, **36**(1), 71–75.
44. Rogers, T. D., McKimm, E., Dickson, P. E., Goldowitz, D., Blaha, C. D., & Mittleman, G. (2013). Is autism a disease of the cerebellum? An integration of clinical and pre-clinical research. *Frontiers in systems neuroscience*, **7**, 15.
45. Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, **43**(10), 1391–1399.
46. Schlerf, J., Ivry, R. B. & Diedrichsen, J. Encoding of Sensory Prediction Errors in the Human Cerebellum. *J. Neurosci.* **32**, 4913–4922 (2012).
47. Schmahmann, J. D. From Movement to Thought : Anatomic Substrates of the Cerebellar Contribution to Cognitive Processing. **198**, (1996).

48. Schmahmann, J. D., & Pandya, D. N. (1997). Anatomic organization of the basilar pontine projections from prefrontal cortices in rhesus monkey. *Journal of Neuroscience*, 17(1), 438-458.
49. Schmahmann, J. D. (1998). Dysmetria of thought: clinical consequences of cerebellar dysfunction on cognition and affect. *Trends in Cognitive Sciences*, 2(9), 362-371.
50. Schmahmann, J. D. & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome. *Brain*, 121(4), 561–579.
51. Sereno, M. I., Diedrichsen, J., Tachrount, M., Testa-Silva, G., d'Arceuil, H., & De Zeeuw, C. (2020). The human cerebellum has almost 80% of the surface area of the neocortex. *Proceedings of the National Academy of Sciences*, 117(32), 19538-19543.
52. 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.
53. 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.
54. Stoodley, C. J. & Schmahmann, J. D. Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *Neuroimage* 44, 489–501 (2009).
55. Strick, P. L., Dum, R. P. & Fiez, J. A. Cerebellum and nonmotor function. *Annu Rev Neurosci* 32, 413–434 (2009).
56. van Es, D. M., van der Zwaag, W., & Knapen, T. (2019). Topographic maps of visual space in the human cerebellum. *Current Biology*, 29(10), 1689-1694.
57. Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. *Neuroimage*, 86, 554-572.
58. Von Bartheld, C. S., Bahney, J., & Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *Journal of Comparative Neurology*, 524(18), 3865-3895.
59. Voogd, J., & Glickstein, M. (1998). The anatomy of the cerebellum. *Trends in cognitive sciences*, 2(9), 307-313.
60. White, J. J., & Sillitoe, R. V. (2013). Development of the cerebellum: from gene expression patterns to circuit maps. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(1), 149-164.
61. Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880-1882.
62. Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9), 338-347.
63. Walters, J., King, M., Bissett, P.G., Ivry, R.B., Diedrichsen, J., Poldrack, R.A. Predicting brain activation maps for arbitrary tasks with cognitive encoding models (under review at *Journal of Cognitive Neuroscience*, 2022).
64. Zhi, D., King, M., Hernandez-Castillo, C. R., & Diedrichsen, J. (2021). Evaluating brain parcellations using the distance-controlled boundary coefficient. *Human Brain Mapping*.