Neural processes underlying cognitive control during language production

Tara Pirnia

June 5, 2024

Joint PNC-MLD program
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Leila Wehbe *chair*
Nazbanou Nozari
Graham Neubig
Stephanie Ries, San Diego State University
Paula Clemens, MD

Abstract

Language production requires the coordination of multiple cognitive processes, including the selection of appropriate words to convey intended meanings and the suppression of competing or irrelevant information. These processes rely on cognitive control mechanisms that enable speakers to successfully navigate the demands of communication. Investigating the neural underpinnings of cognitive control in language production is crucial for understanding both typical and atypical language function, as impairments in these mechanisms are associated with various language disorders, such as aphasia. This thesis aims to elucidate the neural dynamics of cognitive control during word production by employing a combination of experimental paradigms, electrophysiological recordings, and machine learning techniques. We focus on two aspects of cognitive control: (a) the suppression of prepotent word representations (Stroop-like and Picture-word interference tasks) and (b) the resolution of competition arising from the co-activation of contextually-related word representations (blocked-cyclic picture-naming task). By characterizing the spatiotemporal signatures of these processes and their generalizability across individuals, our analysis works to elucidate mechanisms theories of language production and cognitive control. Furthermore, we investigate the extent to which the neural mechanisms underlying cognitive control in language production overlap with those in non-linguistic domains providing insights into the domain-generality of these mechanisms.

Contents

List of Figures

Chapter 1

Introduction

The ability to communicate effectively through language is a hallmark of human cognition. Central to this ability is the process of language production, which involves the rapid and precise coordination of multiple cognitive processes, from the conceptualization of an intended message to the articulation of the corresponding words (Levelt et al., 1999). A key challenge in language production is the selection of appropriate words from a vast mental lexicon in the face of competition from semantically or phonologically related alternatives (Dell, 1986). This selection process engages cognitive control mechanisms to resolve competition and ensure that the intended word is produced (Nozari et al., 2016).

At each point in time, a speaker has various options for what to produce and how to produce it. For example, you can say "I enjoyed the food" or "The pasta was delicious." Owing to years of research in language production, we know that words compete for production, especially if they are similar, e.g., when you say "pasta," "panini" competes for production, and if competition is not resolved successfully, production ends in an error. It is often thought that speakers resolve such competition by applying inhibitory control, a process whereby the competitors are suppressed in favor of the target word. But what is the nature of this control? Is it using the same process that speakers apply when they intentionally wish to suppress an utterance that is on their mind? Imagine the pasta tasted awful, but you wished to be polite. In that situation, you willfully apply cognitive control to override the production of the prepotent word "awful" in favor of "good". While we know that speakers can do both, i.e., suppress prepotent words and resolve competition induced by similar words, we know little about the nature of the control processes underlying these two situations.

Cognitive control specifically refers to the ability to flexibly adapt behavior in accordance with

goals and to suppress prepotent or irrelevant information (Miller& Cohen, 2001). In the context of language production, cognitive control is crucial for overriding automatic or habitual responses in favor of task-relevant responses. For instance, when asked to name the color of the ink in which a word is written, as in the Stroop task (Stroop, 1935), individuals must suppress the prepotent response of reading the word and instead name the ink color. Similarly, when naming pictures in the presence of semantically or phonologically related distractors, as in the picture-word interference paradigm (Schriefers et al., 1990), individuals must resolve the competition between the target and distractor words to produce the intended response.

Broadly, this work looks at how the stages of word production interact in time and space to inform existing theoretical frameworks of language production. We aim to advance theories of cognitive control to characterize the underlying neural processes using machine-learning techniques, applied to EEG and functional neuroimaging recorded during well-established experimental paradigms. Specifically, we characterize the neural features within and across participants, between specified language processes, and in addition, between language and non-language control mechanisms during high and low-conflict conditions. We examine the temporal dynamics and spatial location by systematic mapping of the neural activity to psycholinguistic variables to delineate the relationship between stages of word-production.

From a clinical perspective, impairments in cognitive control have been implicated in various language disorders, such as aphasia (Novick et al., 2005), specific language impairment (Im-Bolter et al., 2006), and developmental dyslexia (Altemeier et al., 2008). Individuals with these disorders often exhibit difficulties in tasks that require the suppression of prepotent responses or the resolution of competition, suggesting that deficits in cognitive control may contribute to their language impairments. By elucidating the neural bases of cognitive control in language production, we can better understand the nature of these impairments and inform the development of targeted interventions.

1.1 Specific Aims

Aim I. Investigate the neural signatures and temporal dynamics of prepotent response suppression during conflict resolution (Chapter 4).

We examine cognitive control in language production during a Stroop-like task. Such tasks require the suppression of prepotent words via cognitive control to successfully accomplish the task's goals. First, we tested the feasibility of using statistical learning methods to study the neural processes underlying this phenomenon from electrophysiological recordings. Once estab-

lished, we implement these tools to characterize the temporal dynamics of task-related control processes. Subsequently, we ask if there is something more general about these states that is common to all humans by examining the temporal variability and generalizable neural features between study participants.

Aim II.   To understand more broadly the neural processes of control within language, we study how conflict is induced in different language production processes. Here, we look to conflict states resulting from contextual similarity. We first characterize the temporal dynamics of high- and low-conflict states due to the co-activation of semantic and phonologically similar word representations by implementing the methods described in Aim I. Next, we focus on the degree to which high-conflict states are generalizable within language production. Specifically, we address (1) whether the observed neural signatures of conflict in semantically similar tasks are generalizable to those of phonologically similar conditions (2) if the features used for decoding in contextual similarity conditions are generalizable to conflict states decoded during Stroop-like tasks through a series of cross-classification procedures.

Aim III.   We examine the generality or specificity of high versus low conflict states between linguistic and non-linguistic domains to understand how the control recruited during language production aligns with the broader control literature. We present a second experimental data set in which participants complete two interleaved experimental tasks that elicit control in language production and visual-motor domains independently. The proposed work leverages the high spatial resolution of a different neuroimaging modality, fMRI recordings, to answer whether and to what extent are the neural correlates of control shared in language and visual-motor domains. We describe the planned whole-brain and searchlight analysis to localize control within and between the two domains.

The present thesis aims to advance our understanding of the neural mechanisms underlying cognitive control in language production through a series of experimental studies and analyses. In Chapter 2, we provide a comprehensive review of the current literature on cognitive control, language production, and their neural bases, highlighting the key theoretical frameworks and empirical findings that inform our research. Chapter 3 describes the methodological approaches and experimental paradigms employed in this thesis, including the Stroop-like task, picture-word interference paradigm, and the interleaved language production and visual-motor task, along with the statistical learning methods applied to electrophysiological and functional neuroimaging data.

The work of this thesis is organized around the three specific aims. Chapter 4 addresses Aim I by investigating the neural signatures and temporal dynamics of prepotent response suppression dur-

ing conflict resolution in a Stroop-like task, utilizing statistical learning methods to characterize task-related control processes and their generalizability across participants. Chapter 5 focuses on Aim II, examining the neural processes of control within language production by studying conflict induced by semantic and phonological similarity. This chapter explores the temporal dynamics and generalizability of high- and low-conflict states within language production, as well as their relation to conflict states decoded during Stroop-like tasks. Chapter 6 tackles Aim III, leveraging the high spatial resolution of fMRI to investigate the generality or specificity of high versus low conflict states between linguistic and non-linguistic domains. Through whole-brain and searchlight analyses, this chapter aims to localize control within and between language production and visual-motor domains.

Finally, Chapter 7 discusses implications of the findings from the experimental studies for theories of cognitive control and language production. This concluding chapter also outlines the limitations of the current work and suggests potential avenues for future research, emphasizing the importance of understanding the neural bases of cognitive control in language production for both basic science and clinical applications.

Chapter 2

Background

2.1   Overview

Language production involves the coordination of multiple processes from the conceptualization of an intended message to the articulation of the corresponding words. The efficiency and accuracy of this process depend on the successful recruitment of cognitive control mechanisms, which enable flexible adaptation of behavior across various cognitive domains. In the context of language production, cognitive control plays a crucial role in resolving competition during word selection and suppressing inappropriate but more likely word choices, particularly in single-word production tasks. This chapter aims to provide a concise overview of the key concepts related to these control processes, laying the foundation for the experimental work and analyses presented in the subsequent chapters of this thesis.

2.2   Cognitive control

Cognitive control refers to the set of processes that allow individuals to flexibly adapt their behavior in accordance with a goal or task, particularly in the face of competing or conflicting information (Botvinick et al., 2001). It involves the ability to maintain and update goal-relevant information, inhibit prepotent or irrelevant responses, and switch between tasks or mental sets. Cognitive control is essential for goal-directed behavior and is particularly important in novel or challenging situations where automatic or habitual responses are insufficient. When the likelihood of successfully achieving a goal is low, greater effort is needed to reach the goal, resulting in longer reaction times and higher error rates (Ridderinkhof et al., 2004).

2.3   Language Production

Language production is composed of the conceptualization of an intended meaning to produce, the subsequent formulation to transform the concept into words, and the final transformation of words to articulation through coordinated muscular movements to produce words (Levelt et al., 1999). Centrally, producing words requires the mapping of concepts to specific lexical representations as well as the mapping of lexical features to their corresponding phonological components (Nozari & Dell, 2009). Word production can be broken down into conceptual preparation, lexical selection, phonological code retrieval, and articulation (Levelt et al., 1999). However, the precise nature of these language production processes in the brain remains unsettled (Nozari & Pinet, 2020).

Several models for such processes have been proposed based on evidence from behavioral and neurological studies. Key models include Levelt's strictly serial model (Levelt et al., 1999) that proposes a sequential, feed-forward mechanism where each stage occurs independently and in order. Lexical access involves two steps: lemma retrieval, which corresponds to the selection of a word's syntactic and semantic properties, and word-form encoding, which involves the retrieval of the word's phonological form. In the context of picture-naming, the timeline of Levelt's serial model delineates the peak of each stage to occur as follows (Indefrey & Levelt, 2004):

- Conceptual preparation: at 175ms

- Lexical selection: between 175-250ms

- Phonological code retrieval: at 355ms

- Articulation: at 455ms post picture onset

In contrast to the strictly serial model, cascading and interactive models of word production propose that information from one stage can influence processing at subsequent stages before the first stage is complete (Dell et al., 2014). These models suggest that there is a continuous flow of information between the semantic, lexical, and phonological levels of processing, and that activation at one level can cascade to influence activation at subsequent levels. Interactive models further propose bidirectional influences between levels, such that activation at a later stage can feed back to influence processing at an earlier stage (Dell, 1986; Nozari & Dell, 2009). Evidence for cascading and interactive processing in word production comes from studies showing that factors such as semantic and phonological similarity can influence both the speed and accuracy of word production (Nozari & Pinet, 2020).

Within the stages of word production, the selection process may require the implementation of

18

cognitive control, although the exact nature of this control is debated (Nozari & Pinet, 2020). Nozari et al. (2016) distinguish between two types of cognitive control in word production: selection control and post-monitoring control. Selection control refers to the process of resolving competition between co-activated word representations during lexical selection, while post-monitoring control involves the detection and correction of errors after a word has been selected but before it is articulated.

While selection control may be specific to the language production system, post-monitoring control may recruit a domain-general mechanism that is shared with other cognitive control tasks. The extent to which these two types of control rely on shared or distinct neural mechanisms remains an open question. More recently, Pinet and Nozari (2023) found that the electrophysiological signatures of Stroop-like interference and contextual similarity interference were distinct, suggesting that the two types of control may rely on different neural mechanisms. Specifically, Stroop-like interference showed a biphasic pattern in representational similarity analysis (RSA), with an initial increase in similarity followed by a decrease, while contextual similarity interference showed a sustained increase in similarity throughout the production process (Pinet & Nozari, 2023). These results suggest distinct mechanisms underlying selection control and post-monitoring control.

2.4   Language disorders and cognitive control

Models of language production have significant implications for understanding and treating neurological disorders that affect language. Cascade and interaction models, which describe information flow and processing interactions in the brain during word production, provide a mechanistic framework for how the human brain produces words (Dell et al., 2014; Nozari & Pinet, 2020). These models can inform diagnosis, treatment strategies, and supportive therapies for language impairments. Conversely, studying the clinical presentation and assessment of these impairments allows for the development of more expressive models. This bidirectional approach is particularly important in language and language disorders research, as language is considered a uniquely human cognitive function, making animal experimentation less applicable for advancing our understanding. Current therapies for language recovery after stroke or traumatic brain injury are notably less effective compared to motor control recovery, despite extensive rehabilitation efforts (Nozari & Pinet, 2020).

2.4.1 Aphasia

Aphasia, a language disorder usually caused by damage to the left cerebral cortex due to stroke, traumatic brain injury, or neurodegenerative disease, is typically categorized based on verbal fluency, comprehension, and repetition abilities (Cahana-Amitay & Albert, 2015). The specific symptoms and severity of aphasia depend on the location and extent of brain damage. Cascade and interaction models can help explain the patterns of impairment observed in different types of aphasia. For example, in Wernicke's aphasia, characterized by fluent but meaningless speech with neologisms, there may be a disruption in the flow of information from the semantic to the phonological level (Schwartz et al., 2006). In contrast, Broca's aphasia, characterized by non-fluent, halting speech and agrammatism, may involve a disruption in the interaction between syntactic and phonological processes (Hula & McNeil, 2008). Current aphasia therapies focus on specific semantic, syntactic, and/or phonological rehabilitation, but the study of their interactions and implications for rehabilitation is still in its early stages (Nozari & Pinet, 2020).

Progressive aphasia, associated with different types of dementia, is characterized by a gradual loss of language abilities. Serial models have been used to characterize language deficits in primary progressive aphasia (PPA) variants, such as semantic and logopenic PPA (Martínez-Cuitiño & Barreyro, 2020). These models predict that semantic variant PPA affects downstream processes like word retrieval and comprehension due to the deterioration of semantic knowledge (Montembeault et al., 2019), while logopenic variant PPA disrupts the mapping between phonological and articulatory levels due to phonological processing deficits (Petroi et al., 2021). Alzheimer's disease (AD), the most common type of dementia, also progressively impairs language. Cascade and interaction models suggest that the deterioration of semantic knowledge in AD affects downstream phonological processes, leading to naming difficulties and circumlocutions (Balthazar et al., 2008; Melrose et al., 2009). These insights can inform the design of cognitive interventions that focus on strengthening the connections between semantic and phonological representations to maintain language function in AD.

2.4.2 Developmental language disorders

Developmental language disorders, such as specific language impairment (SLI) and dyslexia, are characterized by difficulties in language acquisition and processing without other cognitive or sensory impairments. Cascade and interaction models provide insights into the underlying mechanisms of these disorders. For example, children with SLI may have phonological processing deficits that disrupt the interaction between phonological and syntactic processes during word production (Lahey & Edwards, 1999). Similarly, dyslexia has been associated with phonological

processing deficits that affect mapping between orthographic and phonological representations (Ramus & Szenkovits, 2008). Understanding these mechanisms can inform the development of educational interventions and accommodations for children with language learning difficulties.

2.5   Summary

Studying the applicability of these models to neuroimaging studies of word production is crucial for understanding the nature of language impairments in various neurological diseases and disorders. These models provide a framework for how information flows and interacts across different processing levels, guiding the development of targeted diagnostic tools, prognostic markers, and treatment strategies (Nozari & Pinet, 2020). The experimental work presented in this thesis aims to elucidate the neural dynamics of cognitive control during word production by investigating the temporal and spatial characteristics of conflict resolution processes. Specifically, we focus on the suppression of prepotent word representations (Aim I), the resolution of competition arising from contextually-related word representations (Aim II), and the domain-generality of cognitive control mechanisms between linguistic and non-linguistic domains (Aim III). By characterizing the neural signatures of these processes and their generalizability across individuals, we seek to advance theories of language production and cognitive control, informing our understanding of both typical and atypical language function.

Chapter 3

Methods

3.1 Motivation and Approach

This thesis characterizes the underlying neural processes by leveraging statistical learning methods, neural electrophysiology, and neuroimaging recordings in conjunction with experimental manipulations while participants complete each task. We utilized analysis methods that characterize the shared neural features across participants, between specified language processes, and in addition, between language and visuospatial domains. Overall, we aimed to (1) advance theories of cognitive control by the systematic mapping of neural activity to psycholinguistic variables with established experimental paradigms shown to elicit control; (2) test the central assumptions imposed by theoretical frameworks of word production by examining the temporal dynamics of stages of production.

3.2 Empirical Data

Data used in our analyses were obtained from the study by Pinet and Nozari (2023) and from an yet unpublished study by the Nozari lab. We describe below the empirical setup of these studies.

3.2.1 Picture-Stroop paradigm

Participants completed a blocked-cyclic picture-naming task where they named a small set of pictures that are presented in blocks of repeated picture pairs (Belke et al., 2005; Schnur et al., 2009; Nozari et al., 2016). Across the paradigm, a six-way cross with two Stroop-like conditions and three contextual conditions. The Picture-Stroop task condition, a variant of the classic Stroop

task (Stroop, 1935), is designed to measure cognitive control in language production. Stroop-like conditions consisted of half-blocks where participants saw one of two pictures (e.g., cake/pie) and either named them by their canonical names (congruent naming trials) or by the name of the other picture (incongruent naming trials).

The three contextual conditions within the blocked-cyclic picture-naming paradigm were presented block-wise, where picture pairs presented in each block were either semantically related (e.g., cake/pie), phonologically related (e.g., tie/pie), or unrelated (e.g., tie/rake). In total, ten unique pictures were presented in an interleaved fashion, with ten repetitions of each (Figure 3.1).

(a)   (b)
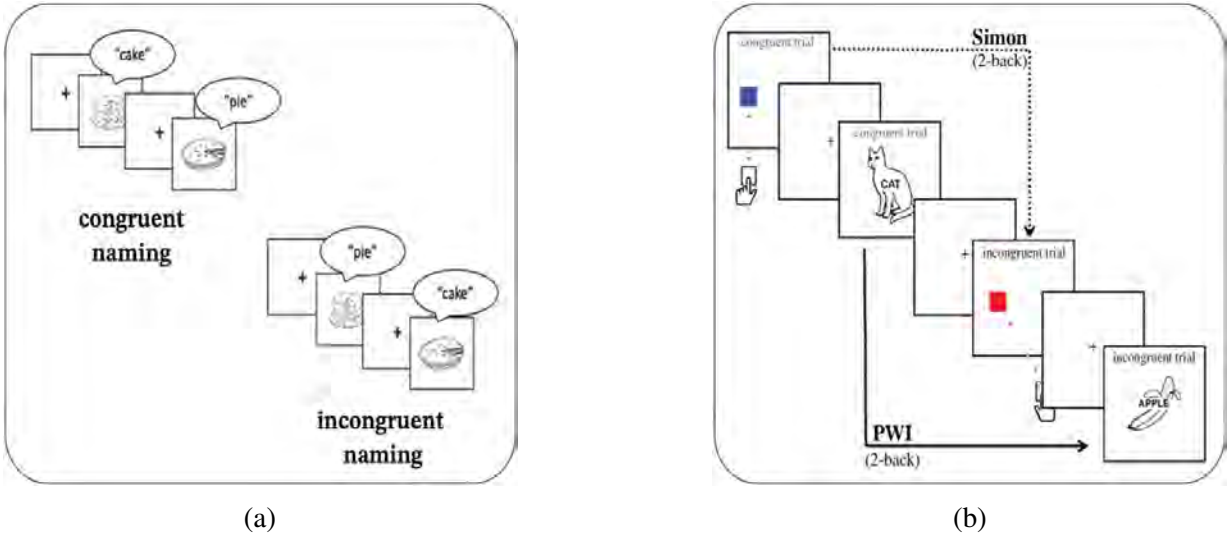
Figure 3.1: Experimental Paradigms. (a) Blocked-cycling picture naming task, and (b) Picture-word interference and Simon task.

3.2.2   The Congruency Sequence Effect Paradigm

To examine cognitive control in the language domain as compared to other cognitive domains, and identify unique and generalizable brain regions, we utilized an experimental paradigm from Freund and Nozari (2018) that combines the picture-word interference (PWI) task and the Simon task. This paradigm allows for the investigation of cognitive control in both the language production and visuospatial domains.

The picture-word interference (PWI) task is a widely used paradigm for studying cognitive control during language production (Schriefers et al., 1990). In this task, participants are presented with a line drawing (picture) and a superimposed word, and are asked to name the picture (target). The task includes congruent trials, where the superimposed word corresponds with the picture

name (e.g., the word "cat" over a picture of a cat), and incongruent trials, where the superimposed word does not match the picture name (e.g., the word "dog" over a picture of a cat) and acts as a distractor (Freund et al, 2016; Shitova, et al, 2017). Previous findings have shown that in the presence of distractors, speakers have longer reaction times and increased errors, as compared to when no distractors are presented (de Zubicaray et al., 2001, de Zubicaray & McMahon, 2009, Piai et al., 2013). The effect of distractors within the PWI task is further modified by the relationship between the target and distractor items (e.g., items related in semantics such as "pie" and "cake"; Schriefers et al., 1990, Alario et al., 2000, Costa et al., 2005).

The Simon task is a non-linguistic visuospatial task that assesses cognitive control in the visuospatial domain (Simon & Rudell, 1967). In the version of the Simon task used in this experiment, participants are required to respond to the color of a stimulus (red or blue) presented on the left or right side of the screen using left and right button presses. The task includes congruent trials, where the stimulus color and its location correspond (e.g., red stimulus on the left side requiring a left button press), and incongruent trials, where the stimulus color and its location do not correspond (e.g., red stimulus on the right side requiring a left button press). The Simon task has been used in fMRI studies to investigate the neural correlates of cognitive control in the visuospatial domain (Liu et al., 2004; Maclin et al., 2001).

In the experiment described, the PWI and Simon tasks were interleaved in a specific design to examine cognitive control across domains (language and visuospatial) and within tasks (Freund & Nozari, 2018). The tasks were structured such that every trial switched between the PWI and Simon tasks, allowing for a 1-back examination of cognitive control across domains and a 2-back examination within tasks. Both tasks included congruent and incongruent trials, which increased conflict during incongruent trials and required cognitive control to resolve that conflict (Hirschfeld et al., 2008; Schriefers et al., 1990). This experimental design allows for a comprehensive investigation of the neural mechanisms underlying cognitive control across and within different cognitive domains.

3.2.3 Electroencephalography

Aims I and II examine the neural activity recorded with electroencephalography (EEG), a non-invasive technique that measures electrical signals generated by neurons from electrodes attached to the scalp (Berger, 1929). Notably, EEG provides high-temporal resolution (milliseconds) measures of the dynamic neural signatures that are associated with the brain states of participants as they complete each trial (Luck, 2014). The following work utilized EEG data collected by Pinet & Nozari (2023) and is composed of a 128-channel EEG recorded as participants completed

the Picture-Stroop task with analysis completed on sensor space.

3.2.4 Functional magnetic resonance imaging (fMRI)

For Aim III, we utilize a second modality to record brain activity—functional magnetic resonance imaging (fMRI). fMRI is an imaging technique that measures changes in blood flow in the brain (Ogawa et al, 1990). As participants perform each task, the neurons require greater oxygen-rich blood, causing the magnetic properties detected by fMRI to change (Kwong et al., 2018, Huettel et al., 2014). While fMRI does not have the temporal resolution of EEG, it provides high-resolution spatial information on the changes in brain activity during task performance.

3.3 Analysis of neural activity

In the following work, we utilize machine learning classifiers to better decode the dynamics of neural activity underlying language production. Supervised classifiers provide a map from neural activity (e.g., EEG and fMRI) to particular task conditions that provide information on when and where stimulus-evoked responses occur in the brain (Varoquaux et al., 2017). Importantly, the proposed thesis demonstrates how machine learning classifiers can directly test and, moreover, inform the theoretical frameworks in the study of language production.

Specifically, we utilize four classifiers: linear ridge classifiers, Support Vector Machines (SVM), decision trees (DT), and logistic classifiers. Briefly, each of these provides a different method for classification—linear ridge classifiers use a linear model with regularization to reduce overfitting; SVM allows for a non-linear approach to separate data points; decision trees utilize a tree-like structure to make class decisions; and a logistic classifier uses a logistic function to model the probability of each condition (Hastie et al., 2009). Linear classifiers, such as ridge regression and logistic regression, are often preferred for neuroimaging data due to their interpretability and robustness to overfitting. However, non-linear classifiers like SVM can potentially capture more complex patterns in the data (Misaki et al., 2010). In the work below, we initially test each classifier's performance in decoding each experimental condition from the recorded neural activity and use the best-performing classifier to test subsequent hypotheses.

Cross-validation was implemented by splitting the data sets to ensure the independence of each model's training from its test performance. Specifically, we utilized a nested-cross-validation scheme to split each data set into a train, validation, and test set to train the model, tune the models' hyperparameters, and to evaluate the models' performance, respectively (Cawley & Talbot, 2010).

To measure the localization of temporal information, we used a sliding window analysis. Sliding window classification involves dividing the neural activity during each trial into a series of overlapping windows and then classifying each window's condition (Fyshe et al., 2019; King & Dehaene, 2014). This allows us to learn how information evolves over time by decoding distinct features across the duration of an EEG recording. By examining the generalizability of classifiers at each subsequent window of time, we can identify when pertinent information is observed and for how long it is sustained. In this study, we used a sliding window size of 50 ms with a 10 ms overlap between windows. These parameters were chosen based on previous work demonstrating their effectiveness in capturing the temporal dynamics of EEG data (Smulders et al, 2018).

The performance of each classifier was measured by classification accuracy by comparing the predicted class labels with the actual class labels to calculate the accuracy score. The accuracy score was the ratio of the number of correctly classified instances to the total number of instances in the test data.

Significance was determined using a non-parametric, cluster-wise statistical test that accounts for the relationship between adjacent time windows. For each model, cluster-size distributions under the null hypothesis were acquired through permutations for each subject. After thresholding classification accuracy of the permutation results at 55%, we defined clusters as accuracy values obtained along all possible train and test windows from the temporal generalization matrices. The null distribution was determined by the maximum cluster size at each permutation (Maris & Oostenveld, 2007). Clusters were defined in the observed results as those accuracy results that were above 55% accuracy and at adjacent train and test time windows. Cluster sizes of $p < .05$, determined by the null distribution, were then determined as significant.

An alternative approach to cluster-based permutation testing is to use false discovery rate (FDR) correction to control for multiple comparisons (Benjamini & Hochberg, 1995). FDR correction can be more sensitive than cluster-based methods, but may be less suitable for neuroimaging data with strong spatial or temporal correlations (Stelzer et al., 2013).

Chapter 4

Aim 1

Aim I : Uncovering the neural signatures and temporal dynamics of conflict resolution during language production

4.1   Introduction

Language production is a complex cognitive process that involves the selection and articulation of words to convey intended meanings. In many cases, word selection is a straightforward process, with speakers producing highly predictable and expected words, such as saying "salt and pepper" when asking someone to pass the condiments. However, there are instances when speakers must override a prepotent response and instead select a less expected or less frequently used word, such as saying "salt and mustard" (Nozari & Novick, 2017; Pinet & Nozari, 2023). This ability to override prepotent responses and select the appropriate word for the intended message is a crucial aspect of language production and requires cognitive control (Nozari et al., 2016).

The current study employs a picture-naming paradigm to investigate the neural correlates of cognitive control during language production. By presenting participants with situations where they must either produce a prepotent word (e.g., say "cat" when they see a cat) or suppress the prepotent word in favor of a different word (e.g., say "dog" when they see a cat), we aim to create low-conflict and high-conflict conditions, respectively (Pinet & Nozari 2023). We hypothesize that the cognitive states will differ between these two conditions, with the former being a low-conflict situation as the picture brings the prepotent label to mind, whereas the latter is a high-conflict situation because the evoked label is different from the target word.

To investigate the neural correlates and temporal dynamics of conflict resolution during language production, which is the primary aim of this chapter, we utilized previously collected participants' electroencephalography (EEG) data during the task and use machine learning techniques to address three main questions that together help to uncover key aspects of cognitive control:

1. Can EEG data be utilized to accurately differentiate between high- and low-conflict states in speakers? To answer this question, we compare the performance of three standard classifiers: a support vector machine (SVM), a decision tree (DT), and a regularized ridge classifier.

2. What is the timeline of overriding a prepotent word in favor of a less potent word? Serial theories of word production assume that lexical selection is completed by 200-300 ms after viewing a picture (Indefrey, 2011; Indefrey & Levelt, 2004). Conversely, interactive theories of production predict lasting effects of lexical selection even during further steps of production (Dell, 1986; Nozari et al, 2016; Pinet & Nozari, 2023). To adjudicate between these accounts, we examine the timeline of cognitive control measured by decodability using the optimal of the three classifiers in a sliding-window approach.

3. Finally, we test the generalizability of neural signatures of low vs. high conflict states across speakers. To determine whether conflict could be decoded from individual participants, we train models using data from all other participants. We also investigate how cross-participant classification changes over time compared to within-participant classifier performance.

4.2 Methods

4.2.1 Experimental paradigm

Electroencephalography (EEG) is a non-invasive technique that measures the electrical signals produced by neurons using electrodes placed on the participant's scalp. EEG provides a highly accurate measure of dynamic neural signatures associated with different brain states during each trial, with high temporal resolution (measured in milliseconds) (Luck, 2014).

Our study examines a previously validated data set from Pinet and Nozari (2023), where a picture-naming paradigm manipulates the level of conflict between the picture and the target word. By comparing EEG activity during low-conflict (congruent) and high-conflict (incongruent) production states, we aimed to study the temporal dynamics of cognitive control. Applying machine learning classifiers to the EEG data allowed us to decode the level of conflict on a

trial-by-trial basis (Bode et al., 2019; Hebart & Baker, 2018).

4.2.2 Study Participants

All data used in our analyses were obtained from the study by Pinet and Nozari (2023). The study included 30 individuals who were right-handed, native English speakers, and had normal or corrected-to-normal vision. Informed consent was obtained from participants by Pinet and Nozari (2023) under a protocol approved by the Institutional Review Board of Johns Hopkins School of Medicine, and they were compensated for their participation. Of the 30 participants, two were excluded from our analysis—one due to a history of speech impairments and the other because they did not complete all the required conditions within the task. The remaining 28 participants (13 males and 15 females) were between 18-35 years old, with an average age of 24.1 ± 4.9.

4.2.3 Stroop-like picture naming task

The current study involved a picture-naming task that employed paired picture stimuli. Participants completed a Stroop-like task with congruent low-conflict trials and incongruent high-conflict trials (see Figure 4.1). In the congruent low-conflict condition, participants were instructed to name the picture presented, which was expected to require minimal cognitive control. In contrast, in the incongruent high-conflict condition, participants were required to produce a non-matching word that belonged to the alternate picture, which was designed to require more cognitive control. This condition was intended to activate both the intended response and the presented picture, leading to increased conflict and a condition requiring recruitment of cognitive control.

The present study employed a blocked-cyclic picture-naming task to elicit verbal responses from participants (Schnur et al., 2006). In this task, ten 300 x 300-pixel, black-and-white line drawings were presented to the participants, corresponding to each word to be named. These pictures were presented in blocks of repeated picture pairs, and across the paradigm, a six-way cross with two Stroop-like conditions and three contextual conditions was examined. The Stroop-like conditions, which are the focus of this study, were comprised of half-blocks of trials wherein participants were presented with one of two pictures, such as cake and pie, and instructed to either name the picture by its canonical name (congruent naming trials) or by the name of the other picture (incongruent naming trials). This resulted in a total of ten unique pictures presented in an interleaved fashion, with ten repetitions of each. The stimuli used in this study were retrieved from the Snodgrass database (Snodgrass & Vanderwart, 1980) and Google Images. Previous

research has utilized this paradigm to investigate various aspects of language processing (Belke et al., 2005; Schnur et al., 2009; Nozari et al., 2016).

4.2.4 Decoding Framework

Three supervised classifiers were trained to map from neural activity (e.g., EEG) to task conditions in order to provide information on when and where stimulus-evoked responses occur in the brain. The purpose of using classifiers is to decode and interpret the neural correlates of cognitive processes underlying control during language production (Bode et al., 2019; Hebart & Baker, 2018). For the analysis in this chapter, we implement relatively simple classifiers that have previously demonstrated the successful decoding of EEG data of cognitive tasks (Lotte et al., 2007; Subasi & Gursoy, 2010; Bode et al., 2019). Specifically, we utilize linear ridge classifiers, support vector machines (SVM), and decision trees (DT). Such classifiers were chosen over more complex modeling approaches, such as deep neural networks, for the purpose of interpretability, which, on smaller datasets, can provide insights into underlying neural mechanisms (Hebart & Baker, 2018; Biecek & Burzykowski, 2021). Linear classifiers, such as the ridge classifiers implemented here, are interpretable as the learned weights directly correspond to each feature's importance in the decision function (Haufe et al., 2014). In the case of SVM and DT, visualization of decision boundaries and feature importance allow for interpretability (Subasi & Gursoy, 2010; Breiman et al., 1984).
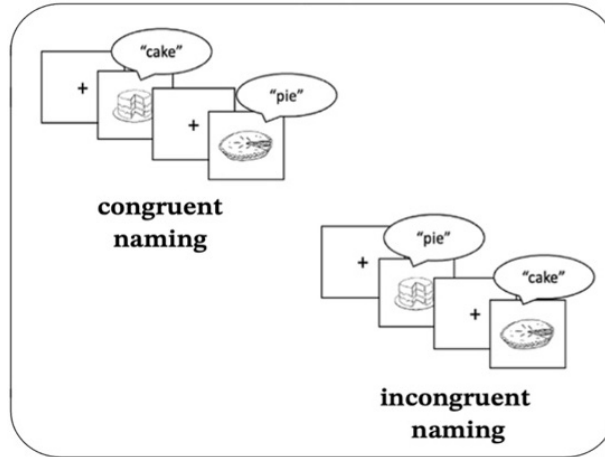
Figure 4.1: Experimental Paradigm: Blocked-cycling picture naming task (Pinet & Nozari. 2023). In the congruent naming example block, participants say "cake" aloud when they are presented with the "cake" picture. In contrast, on congruent naming blocks, participants say "pie" when presented with the "cake" picture. In the latter condition (saying "pie" when presented with a "cake" picture), participants are required to suppress the proponent response (saying "cake") in order to correctly respond to the example stimulus.

Our linear ridge classifiers, SVM, and DT were trained to decode low and high conflict conditions from EEG inputs. For all three models, our input data $\mathbf{X}$, composed of EEG recordings during task completion, with the dimensions N, the number of trials, by C, the number of EEG channels in the trial, by T, the number of time points within a trial and the output, Y, a binary label where the labels are either 'high conflict state' or 'low conflict state'. The input data $\mathbf{X}$ is reshaped, flattening the channel and time dimensions, creating a feature vector for each trial of shape $(N, C \times T)$ with the reshaped data denoted by X (see table 4.1 for notation description).

When using ridge regression for binary classification, a threshold is set on the continuous output of the model. Each input X is assigned to one class, where the model output exceeds the threshold; otherwise, it is assigned to the alternate label class (Hastie et al., 2009). The objective of ridge regression is to find the weight vector w and bias term b that minimize the following cost function:

$$J(w,b) = \frac{1}{2}\sum_{i=1}^{N}\left(y_i - (w^\top x_i + b)\right)^2 + \frac{\lambda}{2}||w||^2,$$

where X_i is the feature vector for the ith trial, obtained from the reshaped input data X_i, $y_i \in -1, 1$ is the transformed binary class label for the trial where -1 corresponds to the 'low conflict state' and 1 corresponds to the 'high conflict state'. The weight vector is denoted by w, the bias term by b and the regularization parameter by λ. The regularization parameter here controls the trade-off between the fitting error and the complexity of the model.

We also examine a more commonly utilized classifier, Support Vector Machines (SVM), which aims to find the hyperplane that maximally separates the two classes in the high-dimensional feature space of the EEG data (Cortes & Vapnik, 1995). The decision function for a binary SVM classifier can be written as:

$$f(x_j) = \text{sign}\left(\sum_{i=1}^{N}\alpha_i y_i K(x_i, x_j) + b\right),$$

where $K(x_i, x_j)$ is the kernel function that measures the similarity between the features training example, i and test example, j.

The third classification method, Decision Trees (DT), is non-parametric. They recursively partition the feature space based on its most informative features, determining the input label by the majority class in the leaf node it falls into (Breiman et al., 1984).

Each of the three classifiers were trained and evaluated on the average of 8 repetitions of EEG recordings spanning the full trial, $X_{full} = [X_{t_0}, \ldots, X_{t_{1000}}]$, where time points, $t = [0ms, 1000ms]$ as well as the lexical selection time window, $X_{lex} = [X_{t_0}, \ldots, X_{t_{350}}]$ (see figure 4.2a; Rifkin & Lippert, 2007; Lotte et al., 2007; Subasi & Gursoy, 2010).

We evaluate the performance of the models using the cross-validation technique, k-fold cross-validation, which involves splitting the dataset into k equal-sized folds with each fold used as a validation set with training the model on the remaining $k - 1$ folds. However, this approach can still suffer from overfitting without properly tuned hyperparameters (Arlot & Celisse, 2010).

To address this issue, a 5-fold nested cross-validation scheme was implemented to split each data set into training, validation, and test sets (Stone, 1974). This technique involves a loop within a loop, where the 'outer loop' reduces the model bias on unseen data and the inner loop is used to choose an optimal hyperparameter (Cawley & Talbot, 2010). The nested cross-validation approach helps to mitigate overfitting to provide a more accurate assessment of the models' ability to generalize as compared to k-fold cross-validation alone (Varma & Simon, 2006).

Model performance was measured by classification accuracy, which compares the predicted class labels with the actual class labels to calculate the accuracy score. The accuracy score is the ratio of the number of instances correctly classified to the total number of instances in the test data (Sokolova & Lapalme, 2009):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

4.2.5 Correlation with muscle artifact

In EEG studies involving tasks with varying levels of difficulty, such as the high- and low-conflict conditions in the current study, there is a concern that the observed differences in neural activity may be confounded by muscle artifacts (Goncharova et al., 2003; Muthukumaraswamy, 2013). These artifacts can arise from increased muscle tension in the face, jaw, or neck during more challenging trials (Porcaro et al., 2019). This is particularly relevant when comparing high-conflict conditions, which may elicit more muscle tension, to low-conflict conditions. To address this concern and ensure that the decoding results are not primarily driven by muscle artifacts, we conducted a participant-specific classification analysis at the sensor level.
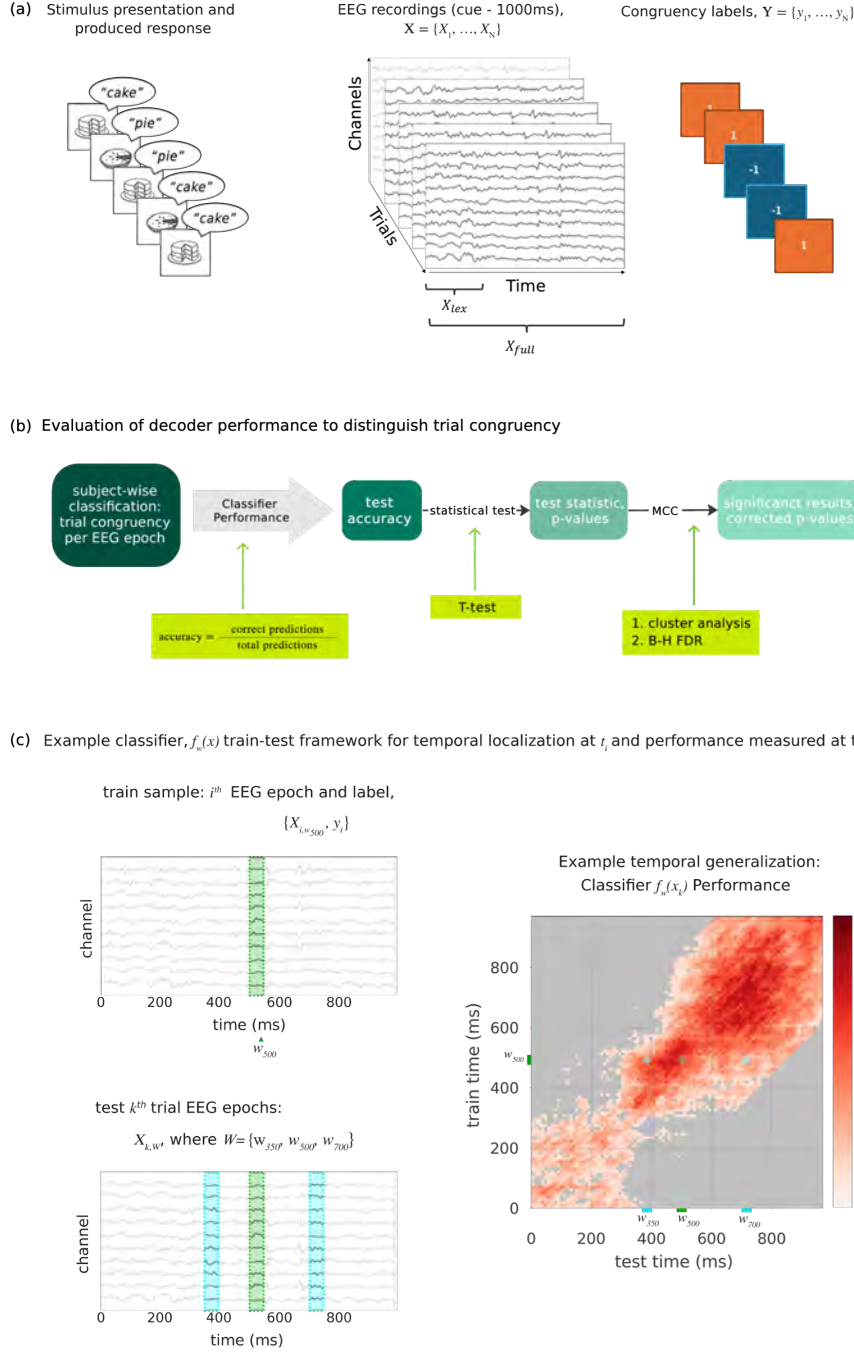
Figure 4.2: Decoding Framework. (a) visualization of stimuli and produced response (left), EEG data recordings per trial (center), and consequent class labels (incongruent (I), congruent (C); right). (b) Evaluation protocol of trained classifier to identify significant clusters of time across temporal generalization results, correcting for multiple comparisons. (c) Visualization of example train and test EEG trial (left) and resultant temporal generalization plot. Top left: Highlighted train sample window is 50ms long, starting at 500 ms after stimulus onset (0ms).

By training and testing classifiers on each participant's EEG channels independently, we can identify which channels consistently decode high- and low-conflict states above chance level. If the above-chance classification accuracy is primarily observed in channels near facial muscles, it would suggest that muscle artifacts may be driving the decoding results. However, if the significant channels are distributed across the scalp, particularly in regions not typically associated with muscle activity, it would provide evidence that the decoding results reflect genuine neural activity related to cognitive control processes (Goncharova et al., 2003; Porcaro et al., 2019).

This analysis helps to rule out the possibility that the observed differences between high- and low-conflict conditions are merely due to muscle artifacts. To eliminate the possibility that facial muscle movement was affecting our analysis, we conducted a participant-specific classification at the sensor level. This involved training and testing a classifier on each participant's EEG channels independently (between 125-128 classifiers per participant). Classifier performance on decoding high- and low-conflict states was tested per participant at each EEG electrode across the full trial period.

A one-sided t-test, ($\mu_0 = 50$), was completed on observed classification accuracies thresholded at 55% (Belke et al., 2005). False positives were controlled using the Benjamini-Hochberg FDR correction with $\alpha = 0.05$ to determine the significance of the classification accuracy of each electrode in all participants (Benjamini and Hochberg, 1995). We can use the Benjamini-Hochberg (BH) correction here instead of the Benjamini-Yekutieli (BY) correction because the data is positively dependent (Genovese et al., 2002).

4.2.6   Classification of temporal dynamics

Temporal localization plays a key role in this study by allowing us to determine the specific time windows in which the brain's response to high and low conflict conditions is most distinguishable. This information is crucial to understanding the temporal unfolding of cognitive control processes and how they relate to the resolution of conflicts (Cohen, 2014; Gratton et al., 2018). The temporal generalization method, as introduced by King and Dehaene (2014), was initially used to investigate the dynamics of neural representations during a task. By training a classifier on data from one time point and testing it on data from other time points, this method reveals the stability and recurrence of neural patterns over time. Similar temporal localization techniques have been applied in previous studies to investigate the timing of neural activity associated with various cognitive tasks (e.g., Fahrenfort et al., 2017; King & Dehaene, 2014; King et al., 2016; Fyshe et al., 2019).

In the context of localizing temporal activity associated with high- and low-conflict conditions, the temporal generalization method can be applied to identify time windows where the neural signatures of these conditions are most distinguishable. By training classifiers on data from different time points and evaluating their performance across the entire trial duration, we can pinpoint the periods when the brain states corresponding to high and low conflict are most separable, thus localizing the temporal dynamics of cognitive control.

To measure the localization of temporal information, a sliding-window analysis was implemented using a ridge classifier trained on 50ms segments with a 10ms step size, on X_{full}. The resulting sets of 50 ms time windows, $X_w = \left[X_{t_w}, \ldots, X_{t_{w+50}} \right]$, where $w = \in \{0, 10, \ldots, 950\}$. Individual classifiers were trained on each of 96 overlapping epochs, f_ws were tested on all epochs, across all time windows $X = \{X_{w_0}, X_{w_{10}}, \ldots, X_{w_{950}}\}$ for all N trials. Figure 4.2c depicts the input EEG and class label of a training instance from the ith trial, starting at time point 500ms. Temporal localization is defined as the test window when the peak test accuracy occurs $X_{peak} = \frac{\#\text{correctly predicted trials}}{\#\text{of total trials}}$.

4.2.7   Participant Cross-classification

To investigate the generalizability of neural signatures across participants, we employed a leave-one-participant-out (LOO) cross-validation scheme (Esterman et al., 2010). In this approach, a classifier is trained on data from all participants except one and then tested on the left-out participant. This process is repeated for each participant, allowing us to assess how well the learned patterns generalize to unseen individuals. The LOO scheme is particularly useful when working with limited sample sizes, as it maximizes the amount of data used for training while still providing an unbiased estimate of the classifier's performance on new participants (Varoquaux et al., 2017).

Temporal localization in the context of the LOO cross-validation scheme involves training classifiers on data from all participants except one at each time point and then evaluating their performance on the left-out participant across all time points. By examining the temporal patterns of classification accuracy, we can identify time windows where the information learned by the classifier is most generalizable across participants. This approach combines the benefits of the temporal generalization method, which reveals the dynamics of neural representations, with the LOO scheme, which assesses the generalizability of these representations across individuals.

For example, if the results show high classification accuracy when trained on data from an early time window (e.g., 200-250 ms) and tested on data from the same window across participants,

this would suggest that the neural signatures of high and low conflict are consistent across individuals during this period. Conversely, if the classifiers trained on later time windows (e.g., 400-500 ms) exhibit poor generalization performance, it may indicate that the neural patterns associated with cognitive control become more variable across participants as the trial progresses.

4.2.8 Statistical analysis

For the sliding window analysis, we implemented a non-parametric, cluster-wise statistical test that accounts for the relationship between adjacent time windows (Maris & Oostenveld, 2007). This permutation test identifies significant differences between conditions by clustering time points while addressing multiple comparisons arising from the large number of tests (Sassenhagen & Draschkow, 2019). The cluster-based permutation test is a statistical method that considers the temporal structure of EEG signals. It assumes that real effects are more likely to occur in large clusters of adjacent time points rather than in small isolated time clusters, and uses an empirical method to determine the extent of the clusters expected to be observed under chance. This test provides a reliable way to determine the significance of classification accuracies across time (Maris & Oostenveld, 2007; Smith & Kutas, 2015).

Specifically, the cluster-based permutation test involves several steps. First, for each time point, a test statistic (e.g., t-value) is calculated between the two conditions being compared. Next, adjacent time points with test statistic values exceeding a predefined threshold (e.g., $p < 0.05$) are grouped into clusters. For each cluster, the sum of the test statistic values within the cluster is calculated, resulting in a cluster-level statistic. The condition labels are then randomly permuted many times (e.g., 1000 permutations), and the previous steps are repeated for each permutation to create a null distribution of cluster-level statistics. Finally, the observed cluster-level statistics are compared against the null distribution, and clusters with statistics exceeding a certain percentile (e.g., 95th percentile) of the null distribution are considered significant. This method effectively controls for multiple comparisons by taking into account the temporal structure of the data and the likelihood of observing clusters of significant differences by chance.

For each model, cluster size distributions under the null hypothesis were acquired by permutations for each subject. We first threshold our results by 55%, which is a rounded value we choose to indicate potential above-chance performance (if we observe a very large cluster where "nodes" have at least 55% performance, and that cluster is much larger than those observed by chance, then we can conclude that the observed region corresponds to higher than chance performance). Other threshold values could be chosen, with a potential tradeoff of sizes of cluster. We chose our value before performing the analysis and did not modify it post-hoc. After thresholding the

DRAFT

classification accuracy of the permutation results at 55%, we defined clusters as accuracy values obtained along all possible train and test windows in the temporal generalization matrices. The null distribution was determined by the maximum cluster size in each permutation (Maris and Oostenveld, 2007). Clusters were defined in the observed results as the accuracy results that were above 55% accuracy and in adjacent time trains and test times. Cluster sizes with $p < 0.05$, determined by the null distribution, were considered significant. Participant-wise distributions of cluster sizes under the null hypothesis were acquired through permutations, with an initial classification accuracy cutoff of 55%.

4.3 Results

4.3.1 Control in high and low-conflict language production tasks

The neural signatures of low and high conflict states are different in the brain and can be robustly decoded via classical ML methods (i.e., phase conditions can be classified from EEG at above-chance accuracy).

For Experiment 1, we conducted a comparison of three classifiers (Ridge, SVM and DM) to detect conflict by decoding congruent and incongruent conditions. Each classifier was trained and evaluated on 8 repetitions of EEG recordings spanning the trial time window from cue to 1000ms. EEG trials and corresponding labels were split into train, validation, and test using a 5-fold nested cross-validation scheme stratified by trial blocks. Accuracy is calculated and averaged per fold across participants. We measured model performance through classification accuracy for each test set. Significance was measured through permutation testing to determine the distribution under the null hypothesis and define significant confidence intervals per model.

Of the three classifiers, ridge and SVM both performed significantly above chance (69.6%, SD=10.6 and 72.6%, SD=10.2, respectively), but DT classifier did not (53.8%, SD=11.6). Moreover, utilizing linear regression, the ridge classifier decoded conditions significantly better than both SVM and Decision Tree (Figure 4.3).

Our results indicate that the neural signatures of conflict, as measured through simple classifiers, can decode congruent and incongruent trials with a ridge regression classifier performing optimally. Overall, the results imply that despite signal noise, the condition effects are robust at a group level and suitable for subsequent analyses.
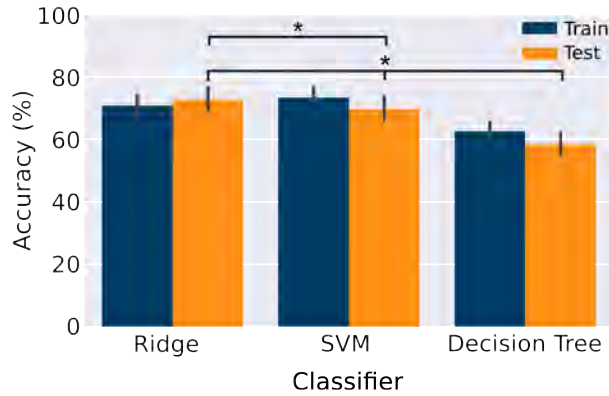
Figure 4.3: Comparison of prediction accuracy from three classifiers: Ridge classifier, SVM, and Decision tree. The training accuracy determined by the inner cross-validation is shown in blue. Test accuracy, based on held-out blocks, is shown in orange. (SVM=support vector machine).

4.3.2 Contribution of facial muscle artifacts

The electrodes with the highest t-values that remain significant after the FDR correction are typically located in the central posterior area. We do not find that electrodes placed around the face are the sole electrodes able to classify above chance (Figure ??). These findings are in opposition to the hypothesis that the above-mentioned chance classification accuracy is caused by electrodes that are linked to an increase in muscle tension during more difficult situations.

The precise timing of cognitive control processes during word production remains a topic of ongoing research. According to Indefrey and Levelt's (2004) model of language production, lexical selection occurs approximately 250-300ms after stimulus presentation. In a strictly serial model (Levelt et al., 1999), lexical selection is thought to eliminate competing lexical representations, preventing their influence on subsequent processing stages. If cognitive control completely resolves lexical conflict during this stage, as stipulated by serial models, we expect the peak classification accuracy to occur within the lexical selection time window.

However, an alternative perspective posits that the language production system operates in a more continuous manner, with lexical competition effects persisting beyond the lexical selection stage (Dell, 1986; Pinet & Nozari, 2023). Under this view, cognitive control processes may be engaged over a longer time course, extending beyond the typical lexical selection window. Consequently, we would expect improved classification accuracy when considering a larger time window that encompasses both lexical selection and later processing stages.

Classification per channel

per channel t-test

fdr-corrected: $\alpha = 0.05$
range : $56.8 - 67.1\%$

fdr-corrected: $\alpha = 0.05$
range : $3.1 - 9.7$

(a)

(b)
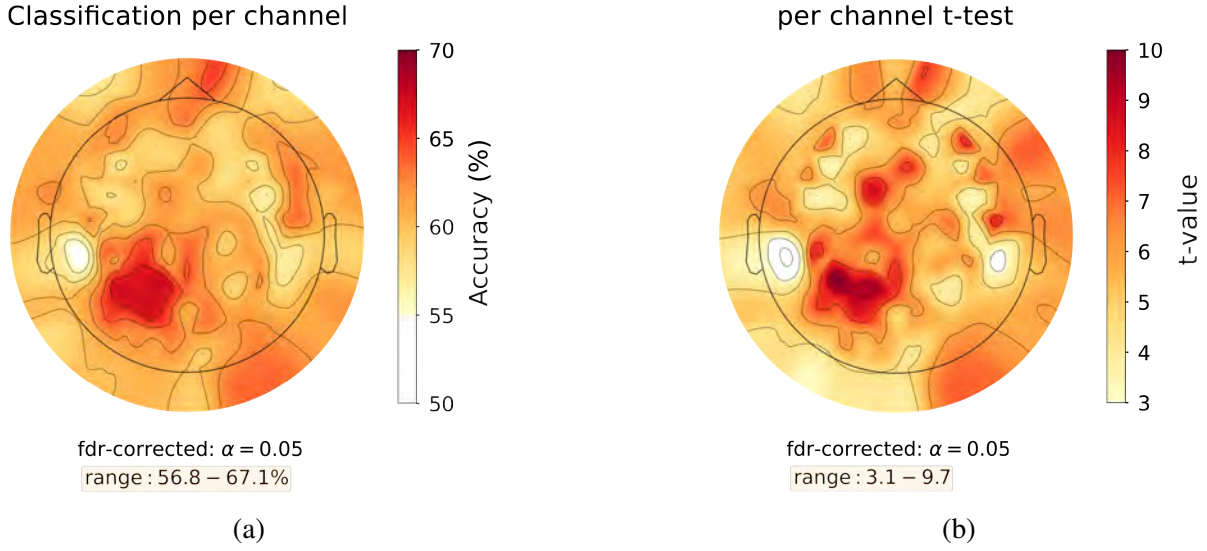
Figure 4.4: Classifier performance per EEG electrode averaged for all participants. (a) The classification accuracy averaged per electrode. (b) Results of the t-test comparing high vs. low conflict classification accuracy. The resulting accuracies were thresholded at 55% and corrected for FDR (= 0.05). The optimal classifier from the full trial analysis (Section 3.5) was selected, trained, and tested on each electrode for each participant.
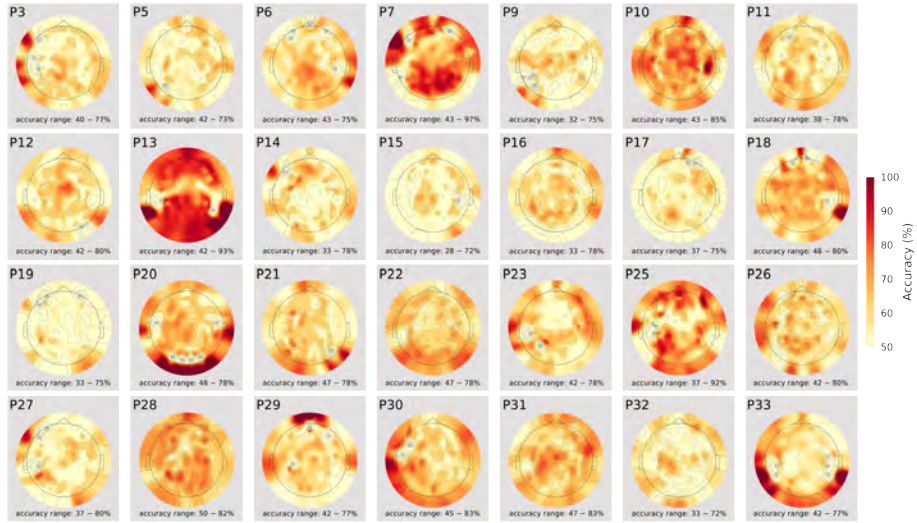
Figure 4.5: Classifier performance per EEG electrode per individual participant. The accuracy of decoding high vs low-conflict classification accuracy. Results for each participant were thresholded at 50% accuracy.

4.3.3 Group-level decoding of conflict states

To adjudicate between these two possibilities, we utilized the methods described in Section 4.3.2 to compare the decodability of cognitive control states within the lexical selection time window

(250-350ms) to that of classifiers trained on the entire trial duration (cue to 1000ms). If cognitive control is indeed limited to the lexical selection stage, as predicted by serial models, classifier performance should not improve when considering a longer time window. Conversely, if control processes extend beyond the theoretical lexical selection window, as suggested by continuous processing models, the inclusion of a larger time window should enhance the decodability of cognitive control states.

We directly compared the performance of ridge classifiers trained and tested on the full trial window (cue to 1000ms) to that of classifiers trained and tested on the expected lexical selection window (cue to 350ms, slightly extended beyond the typical 300ms to account for variability in response times across participants). Our results revealed that the classification accuracy of models trained and tested on the full trial window was significantly higher than that of models trained and tested on the lexical selection window alone (see Figure **??**a). This finding suggests that cognitive control processes likely extend beyond the theoretical lexical selection window, providing support for continuous processing models of language production.
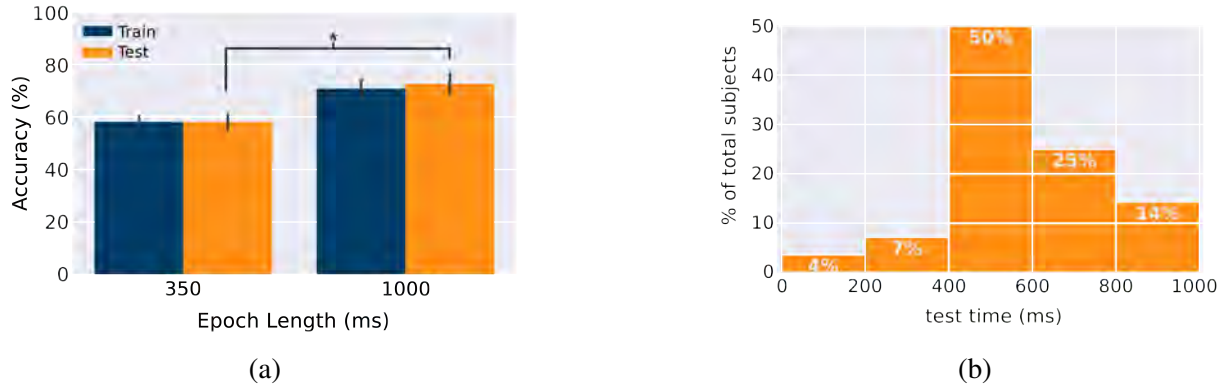
Figure 4.6: Timeline of control. (a) Comparison of ridge classifier performance within an early epoch (350ms) vs. full trial (1000ms). The training accuracy is shown in blue. Test accuracy, based on held-out blocks, is shown in orange. (b) Percent distribution of participants' peak performance train-test time window. Train-test time window where the classifier performs optimally per participant is shown in histogram with 200ms bin width.

4.3.4   Individual timelines in cognitive control

Results presented under Section 4.6.1 suggest the differences between high and low conflict states throughout the trial. There are two possible reasons for this: (a) all speakers show this extended timeline, or (b) different speakers show peak accuracy at different times, resulting in an extended window at the group level. To distinguish between these two possibilities, we examine if the decoded temporal features underlying control are variably localized at the participant level.
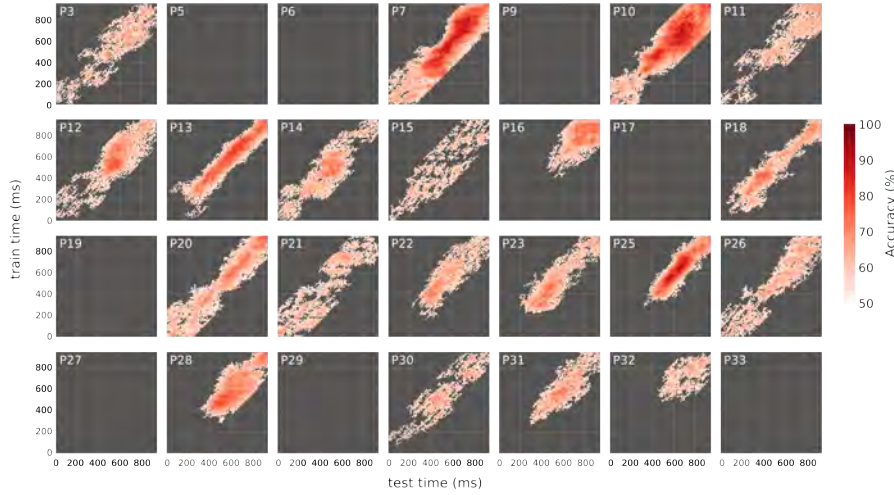
Figure 4.7: Within-participant classification across time. The temporal generalization plots for our 28 subjects of classifier train-test performance averaged on CV folds. Results demonstrate that 20 of them achieved classification accuracy that was higher than what would be expected by chance. y-axis: classifiers' training epoch time window, x-axis: test epoch time windows, color map: performance accuracy clusters threshold at >55% accuracy, corrected for multiple comparisons with FDR correction across train-time windows.

The results of the sliding window analysis revealed that the ridge classifier successfully decoded high and low-conflict conditions, which are thought to engage cognitive control processes, even when given only 50ms of EEG data. Notably, the time windows of peak classification accuracy varied considerably across participants, spanning the entire 1000ms trial duration (Figure 4.6). While some participants exhibited early peak accuracy, the majority showed peak performance after 400ms post-stimulus onset.

The dynamics in classification accuracy across time suggest the resolution of lexical competition, which has been traditionally expected in early time window (200-250 ms), can emerge much later in some speakers. This highlights the critical importance of individual differences in models of language production and cognitive control. The observed variability in peak time windows, reinforced by the findings from Section 4.6.1, underscores the need for further investigation into participant-level differences.

4.3.5 Generalizability of cognitive control across participants

Having demonstrated successful decoding of high and low conflict states within individual participants, we next investigated whether these states have a common neural signature across the population. We trained classifiers on all but one participant and tested on the left-out participant

43

(across-participant classification) and compared the performance to within-participant classification.

(a)

(b)
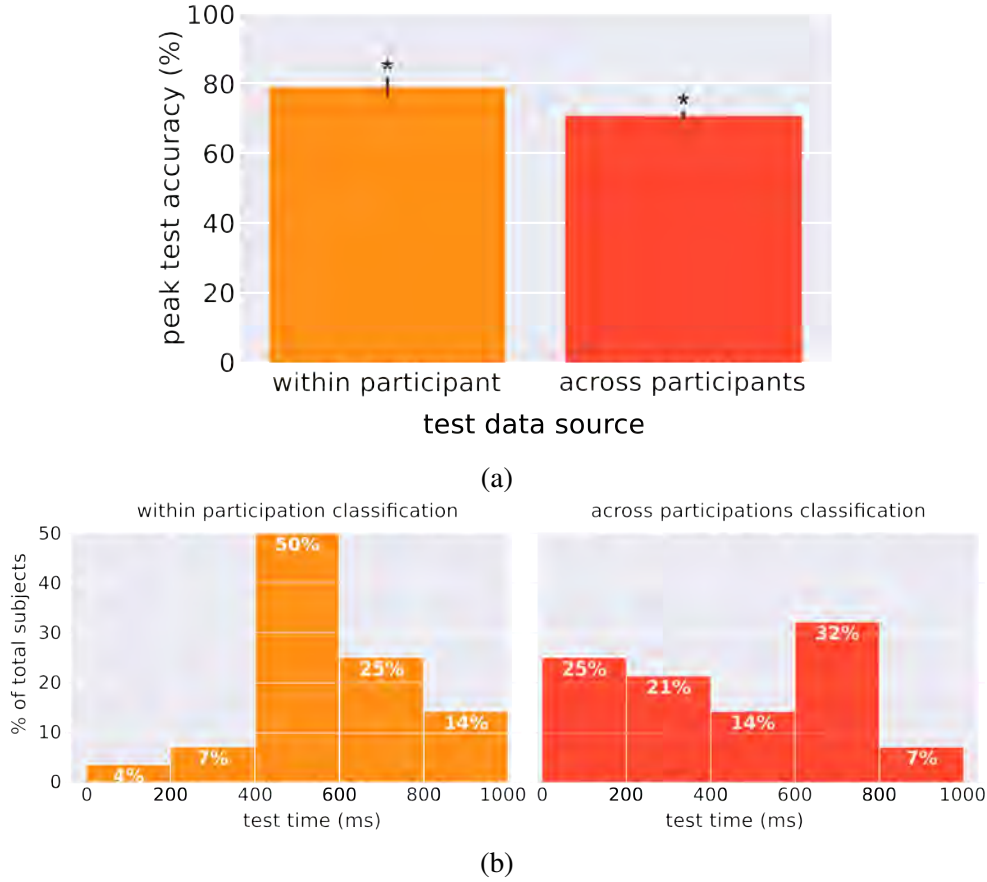
Figure 4.8: Comparison between within and across-participant decoding. (a) Group-level cross-participant decoding: a comparison of peak performance accuracy of within- vs. across-participants trained classifiers. The maximal test accuracy is identified across classifiers trained and tested along an expanded diagonal per participant. (b) Timeline of conflict: Percent distribution of participants' peak performance train-test time window. Train-test time window where the classifier performs optimally per participant is shown in the histogram with 200ms bin width.

Each within-participant classifier is trained and tested on the same participant. Each across-participants classifier is trained on all but a single left-out test participant. To obtain the peak test accuracy value, trials are segmented into 146 overlapping epochs of 50ms in duration and 10ms step size. Individual classifiers were trained on epochs within a single 50ms time window and tested on all possible epochs from the entire trial length.

The results showed that classifiers could decode conflict states above chance accuracy in both within-participant and across-participant settings, but the performance was greater for within-participant classifiers (Figure 4.8a). The temporal distribution of peak performance time win-

dows was similar for both within-participant and across-participant classifiers, with most participants showing peak accuracy after 400ms (Figure 4.8b).

The successful decoding of conflict states using across-participant classifiers suggests the presence of a common neural signature of cognitive control across individuals. However, the higher performance of within-participant classifiers indicates that individual differences in neural signatures also play a role in the decoding of cognitive control states.
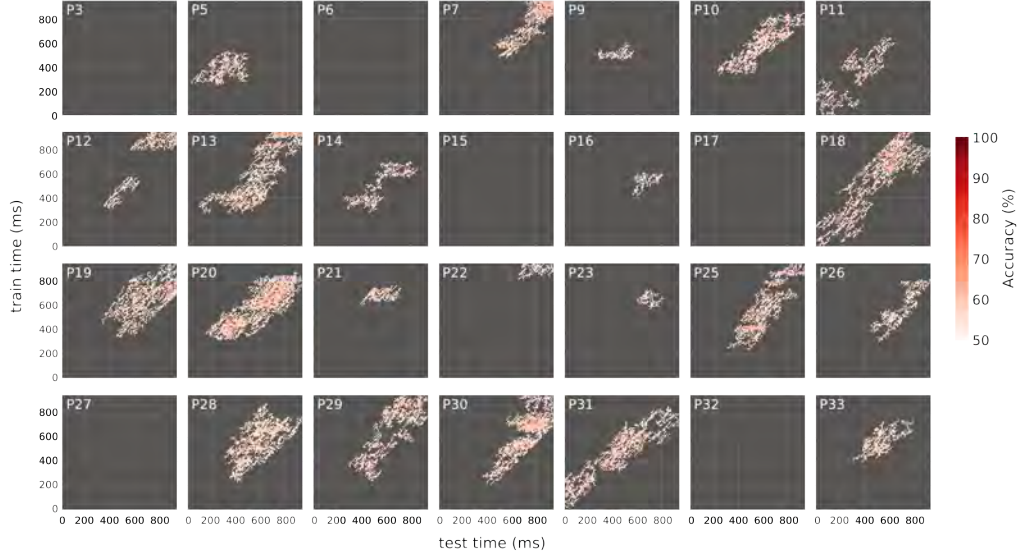
Figure 4.9: Cross-participant classification across time. The temporal generalization plots for our 28 participants (indicated by the number in the top-right corner) of classifiers where each was trained on 27 participants and tested on the left-out participant. The performance shown an average on all CV folds. Results demonstrate that 22 of them achieved classification accuracy that was higher than what would be expected by chance. y-axis: classifiers' training epoch time window, x-axis: test epoch time windows, color map: performance accuracy clusters threshold at >55% accuracy, corrected for multiple comparisons with FDR correction across train-time windows.

To further investigate the generalizability of neural signatures across participants and their temporal dynamics, we conducted a cross-participant temporal generalization analysis, as shown in Figure 4.9. This figure shows the results of the cross-participant temporal generalization analysis. For each participant (indicated by the number in the top-right corner), a classifier was trained on data from the other 27 participants and tested on the left-out participant. The temporal generalization matrices display the classifier performance when trained and tested on different time windows. The results demonstrate that for 22 out of 28 participants, the cross-participant classifiers achieved above-chance accuracy in decoding conflict states. This suggests the presence of

common neural signatures of cognitive control across individuals, although the variability in the temporal patterns also highlights the role of individual difference.

In this chapter, we investigated the neural signatures and temporal dynamics of cognitive control during language production using a Stroop-like picture-naming task and machine learning techniques applied to EEG data. Our results demonstrate that high and low conflict states can be reliably decoded from neural activity patterns, with ridge regression classifiers achieving the highest performance. The successful decoding of conflict states suggests that cognitive control processes elicit distinct neural signatures that can be detected using multivariate analysis methods.

The temporal dynamics of cognitive control were explored using a sliding window analysis, which revealed that the peak classification accuracy varied considerably across participants, spanning the entire 1000ms trial duration. This finding challenges the strictly serial view of language production and suggests that cognitive control processes operate in a more continuous manner, extending beyond the theoretical lexical selection window. The observed variability in peak time windows highlights the importance of considering individual differences in models of language production and cognitive control.

Furthermore, we assessed the generalizability of cognitive control signatures across participants using cross-participant classification. While above-chance decoding accuracy was achieved in both within-participant and across-participant settings, the performance was higher for within-participant classifiers. This result indicates the presence of common neural signatures of cognitive control across individuals, but also underscores the role of individual differences in the neural implementation of these processes. Future research should investigate the factors contributing to these individual differences and their implications for models of language production and cognitive control. Overall, our findings provide new insights into the neural mechanisms underlying cognitive control in language production and highlight the potential of machine learning techniques for uncovering the spatiotemporal dynamics of these processes.

4.4 Discussion

In this chapter, we investigated the neural signatures and temporal dynamics of cognitive control during language production using a Stroop-like picture-naming task and machine learning techniques applied to EEG data. Our results demonstrate that high and low conflict states can be reliably decoded from neural activity patterns, with ridge regression classifiers achieving the highest performance. The successful decoding of conflict states suggests that cognitive control

processes elicit distinct neural signatures that can be detected using multivariate analysis methods.

The temporal dynamics of cognitive control were explored using a sliding window analysis, which revealed that the peak classification accuracy varied considerably across participants, spanning the entire 1000ms trial duration. This finding challenges the strictly serial view of language production and suggests that cognitive control processes operate in a more continuous manner, extending beyond the theoretical lexical selection window. The observed variability in peak time windows highlights the importance of considering individual differences in models of language production and cognitive control.

Furthermore, we assessed the generalizability of cognitive control signatures across participants using cross-participant classification. While above-chance decoding accuracy was achieved in both within-participant and across-participant settings, the performance was higher for within-participant classifiers. This result indicates the presence of common neural signatures of cognitive control across individuals, but also underscores the role of individual differences in the neural implementation of these processes. Future research should investigate the factors contributing to these individual differences and their implications for models of language production and cognitive control. Overall, our findings provide new insights into the neural mechanisms underlying cognitive control in language production and highlight the potential of machine learning techniques for uncovering the spatiotemporal dynamics of these processes.

Chapter 5

Aim 2

Aim II: Decoding the spatiotemporal dynamics of interference due to contextual similarity

5.1 Introduction

As we speak, we elicit processes to map the meaning of what we want to express, the semantic representations, to the specific words we select to convey it—their lexical items. We additionally need to ensure those lexical items are correctly mapped to the sounds or phonological forms we aim to produce. Increases in conflict from various sources can disrupt these underlying processes during language production. One important source of conflict is contextual similarity, where words are generally harder to produce more similar contexts compared to dissimilar contexts (Belke et al., 2005; Nozari et al., 2016; Schnur et al., 2006, 2009). There are two main types of contextual similarity: similarity in meaning, like cat-dog (semantic similarity) or in form, like cat-mat (phonological similarity). Semantic similarity affects the mapping of semantic features to words, while phonological similarity influences the mapping of words to phonemes. Despite their distinct loci, both types of similarity cause interference such as increased naming latencies, through increased conflict with the related competitor word (see Nozari & Pinet, 2020, for a review).

The precise mechanisms underlying lexical selection and the resolution of competition remain a topic of debate. Competitive accounts posit that lexical selection involves direct competition between co-activated representations, with the most highly activated item being selected (Levelt

et al., 1999). In contrast, noncompetitive accounts argue that selection depends on a threshold of activation without direct competition between alternatives (Oppenheim et al., 2010). Psycholinguistic models also differ in their assumptions about the temporal organization of word production processes. Serial models (Levelt et al., 1999) propose discrete stages with unidirectional information flow, implying that semantic-lexical mapping should precede lexical-phonological mapping. Globally modular models (Dell, 1986; Rapp & Goldrick, 2000) allow for bi-directional interaction between levels while maintaining their separation. Non-modular views, in contrast, argue for temporal overlap between semantic and phonological processes (Munding et al., 2016; Strijkers et al., 2017). Investigating the time course of semantic and phonological conflict could provide insights into the temporal coordination of these mappings.

Work by Oppenheim and Nozari (2021) suggests that behavioral interference or facilitation effects alone may not be sufficient to distinguish between these accounts, highlighting the need for neuroimaging and electrophysiological approaches to elucidate the neural dynamics of lexical selection and control. Moreover, Nozari and Pinet (2020) underscore the need for more nuanced investigations of how dynamics of activation within distributed semantic representations can contribute to the observed neural signatures of semantic similarity effects. In line with this perspective, we aim to elucidate the spatiotemporal dynamics of conflict as a function of both semantic and phonological similarity. By applying time-resolved decoding to EEG data, and directly comparing the neural correlates of semantic and phonological similarity, we can test predictions of theoretical accounts positing distinct vs. overlapping time courses for these effects (e.g., Dell, 1986; Levelt et al., 1999).

Previous research findings underscore the utility of electrophysiological methods for investigating the neural dynamics of language production. Costa et al (2009) identified early ERP components (e.g., P2) modulated by lexical competition and Strijkers et al (2017) later identified components (e.g., N400) sensitive to phonological processes. Pinet et al. (2019) revealed the temporal dynamics of word retrieval in high-frequency bands reflect early activation in left temporal regions followed by later frontal activity, further underscoring the utility of electrophysiological methods, and in combination with temporal analysis. However, much remains unknown about the precise spatiotemporal signatures of conflict resolution in word production. Few studies have directly compared the neural correlates of different sources of interference, such as semantic versus phonological similarity (Pinet & Nozari, 2023) or contextual similarity versus Stroop-like conflict (Nozari et al., 2016). Understanding the commonalities and differences between these conflict types could shed light on the domain-generality of cognitive control mechanisms in language production.

The present study aims to address these open questions by leveraging the high temporal resolution of electroencephalography (EEG) and machine learning techniques to decode conflict states due to contextual similarity during language production. Specifically, we ask:

1. Can low and high conflict states induced by semantic and phonological similarity be reliably decoded from EEG signals?

2. Do the neural signatures of semantic and phonological conflict exhibit distinct time courses, as predicted by serial or modular language models?

3. Do different sources of conflict, such as semantic versus phonological similarity (Pinet & Nozari, 2023a) and contextual similarity versus Stroop-like interference (Nozari et al., 2016), engage common neural substrates?
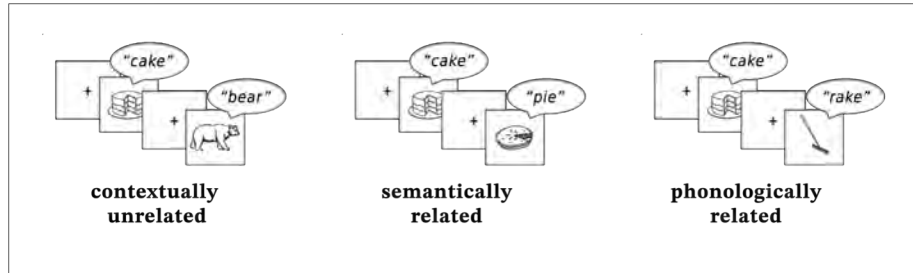
Figure 5.1: Simplified blocked cyclic picture naming task (Pinet and Nozari 2023a).

To address these questions, we examine data acquired by Pinet and Nozari, (2023a), employing a blocked-cyclic naming paradigm (Belke et al., 2005; Schnur et al., 2006), where participants name small sets of pictures that are either semantically related, phonologically related, or unrelated. By utilizing decoding analyses and cross-condition generalization tests to the EEG data, we map out the temporal dynamics and commonality of conflict representations within and across processing levels.

Our findings inform neural models of language production and elucidate the mechanisms of cognitive control over word selection. For example, our second question assists in differentiating serial models (Levelt et al., 1999), as well as globally modular models (Dell, 1986; Rapp & Goldrick, 2000), which posit that semantic-lexical mapping generally precedes lexical-phonological mapping from non-modular models (Strijkers, 2016) which reject such temporal segregation. Collectively, the results of the analyses in this chapter will inform us of the decodability of low and high conflict states induced by semantic and phonological similarity within participants and the timeline of such effects, as well as the generalizability of such states across participants. Additionally, we can investigate whether these conflict states have the same neural

signature when they arise at different stages of processing or through different manipulations. Ultimately, this research improves our understanding of the generalizability of conflict states in language production across participants.

5.2   Methods

5.2.1   Experimental methods

We examine how changes in the relationship of co-activated item representations affect EEG recordings in a related-unrelated picture naming task. This paradigm is a simplified version of the blocked cyclic naming task (Belke et al., 2005; Schnur et al., 2006), where participants repeatedly name small sets of pictures that are either semantically related, phonologically related, or unrelated to each other.

Specifically, our study examined data from Pinet & Nozari (2023a), where participants we[2] re presented with two pictures in each block, with the same pictures repeating multiple times (see Figure 4.1 for details). The motivation for using this task is multifold–The use of two-item blocks where each picture acts as a "distractor" for the other ensures a consistent level of contextual similarity or dissimilarity throughout each block (Nozari et al., 2016; Pinet & Nozari, 2023a). By simplifying the paradigms used by Belke et al. (2005) and Schnur et al. (2006), we reduced confounds related to variations in visual complexity or familiarity differences.[3] It also enhances the signal-to-noise ratio in the EEG recordings by allowing for multiple repetitions of the same stimuli, which is particularly important for time-resolved decoding analyses that require robust patterns of neural activity (Grootswagers et al., 2017; Ries et al., 2021a). Overall, the simplified design enables more distinct identification of the observed effects of linguistic manipulations and facilitates the interpretation of behavioral and neural data (McDonagh et al., 2020).

Compared to other paradigms like picture-word interference (PWI), the blocked cyclic naming task offers several advantages. In PWI, participants name pictures while ignoring super-imposed distractor words, which introduces additional attentional and perceptual demands that can complicate the interpretation of effects (Damian & Martin, 1999; Piai et al., 2012). The blocked cyclic naming paradigm, on the other hand, isolates the processes of interest (i.e., lexical-semantic and lexical-phonological mapping) without the need for active distractor suppression. Moreover, while PWI typically probes the effects of a single exposure to semantic or phonological distractors, blocked cyclic naming has been shown to examine the cumulative interference or

facilitation effects across multiple naming episodes (Biegler et al., 2008; Navarrete et al., 2014).

In summary, the related-unrelated picture naming paradigm employed in this study offers a controlled means to investigate the spatiotemporal dynamics of semantic and phonological processing during language production while minimizing potential confounds and maximizing the sensitivity of behavioral and neural measures. It thus aligns with our aims to elucidate the cognitive and neural mechanisms underlying the resolution of conflict arising from the co-activation of linguistically related representations.

5.2.2 Decoding Framework

To establish that contextually similar vs. unrelated conditions from neural activity recorded with EEG can be distinguished, we applied a similar approach described in Section 4.2. We trained three different classifiers and evaluated their ability to decode the presented conditions. More specifically, we examined how accurately a ridge classifier, a logistic classifier, and an SVM classified each condition in the two sets: semantically related vs. unrelated and phonologically related vs. unrelated.

The splitting of data for cross-validation differed from that in Aim I to address the experimental block structure. In Aim I , we averaged all repetitions of unique picture-target pairs and stratified train-test-validation block-wise, then completed a nested cross-validation (CV) scheme (see Section 4.2.2). Given that the congruent and incongruent names were paired within experimental blocks, block-wise splitting ensures that the picture presented remains independent of the intended target production (i.e. each picture presented with a CV fold appears in both congruent or incongruent conditions). However, the data-splitting scheme from Aim I does not stratify contextually-similarity conditions (i.e., distributing related and unrelated context conditions equally across CV sets) and introduces an interaction between unique picture names and relatedness within the training sets that are inverted in the test set.

To maintain picture-context independence, the cross-validation scheme was changed, allowing for conflict due to contextual-relatedness to be learned. Specifically, three exemplar sets were created, each composed of an average of 2-3 repetitions of all unique picture-target stimuli. The cross-validation scheme was then trained, tested, and validated using an exemplar set, each containing an average of all possible unique trial repetitions, allowing all unique picture stimuli to be independent of contextual similarity. The nested CV approach allowed for hyperparameter tuning and evaluation of the model to remain independent. The resulting classification accuracies were measured as an average of the three possible combinations of train-test-validation splits.

5.2.3 Time-resolved decoding of contextual similarity

Subsequently, the optimal classifier type was then used in a sliding window analysis was completed to evaluate the temporal location of at optimally performing time windows. This analysis required that each full 1000ms trial be segmented into epochs with overlapping time windows of 50ms duration and 10ms step size, respectively. For each participant, 146 independent classifiers are trained, one on each possible epoch (segmented trial data from a single 50ms time window) and tested on all possible unobserved epochs.

Significance was calculated by non-parametric measurement of the null distribution and cluster correction. The null distribution was determined on the performance accuracy of classifiers trained and tested on 500 permutations, threshold at a minimum of 55% accuracy and multiple comparisons based on cluster size and a max $p < 0.05$.
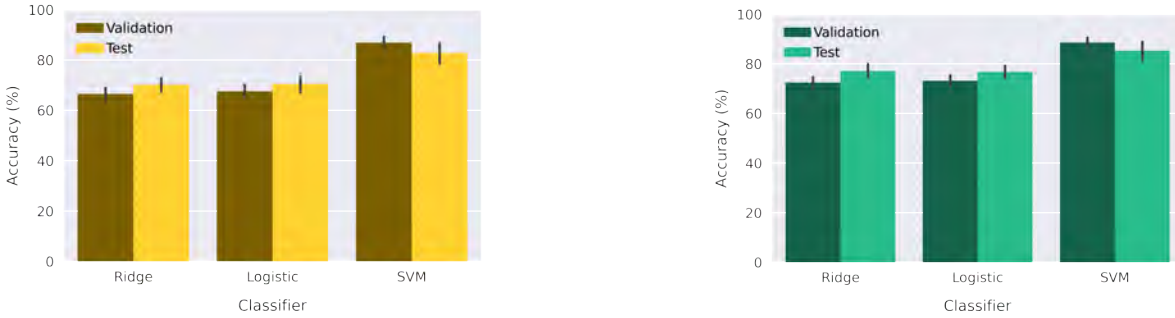
Figure 5.2: Comparison of prediction accuracy from three classifiers (ridge classifier, logistic classifier, and support vector machine) trained and tested on full experimental trials (1000ms) for (a) semantically related and (b) phonologically related interference conditions. The training accuracy determined by the inner cross-validation is shown in blue (support vector machine). The validation and test accuracies are shown in orange and yellow, respectively, for all three classifiers. The support vector machine achieves the highest accuracy in both the semantically and phonologically related conditions, outperforming the ridge and logistic classifiers.

5.3 Results

5.3.1 Differentiating neural signatures elicited by contextual similarity.

In part one of this chapter's aims, we asked if we could successfully decode low and high-conflict states induced by contextual similarity using the same general approach used in the previous chapter to decode these states in a Stroop-like task. The decoding results of semantically related vs unrelated conditions, ridge, logistic regression, and SVM classifiers results showed that

all models performed significantly above chance (70.2%, SD=7.8, 70.4%, SD=8.2, and 81.7%, SD=10.8 respectively). Moreover, the SVM classifier decoded semantically related, high-conflict conditions significantly better than both ridge and logistic regression classifiers (Figure 5.2a). Similarly, in decoding phonologically related and unrelated conditions, the ridge, logistic regression, and SVM classifiers both performed significantly above chance (77.2%, SD=6.9, 76.8%, SD=6.9, and 85.4%, SD=10.4, respectively). In fact, the SVM classifier decoded phonologically related, high-conflict conditions significantly better than both ridge and logistic regression classifiers (Figure 5.2b).

The results in Figure 5.2 address part one of our chapter aim, demonstrating that the high-conflict states, induced by the co-activation of linguistically related stimuli, can be reliably distinguished from the unrelated condition in EEG signals using any of the three classifiers. The superior performance of the support vector machine suggests that it is the most suitable classifier for capturing the neural signatures of conflict in both conditions.
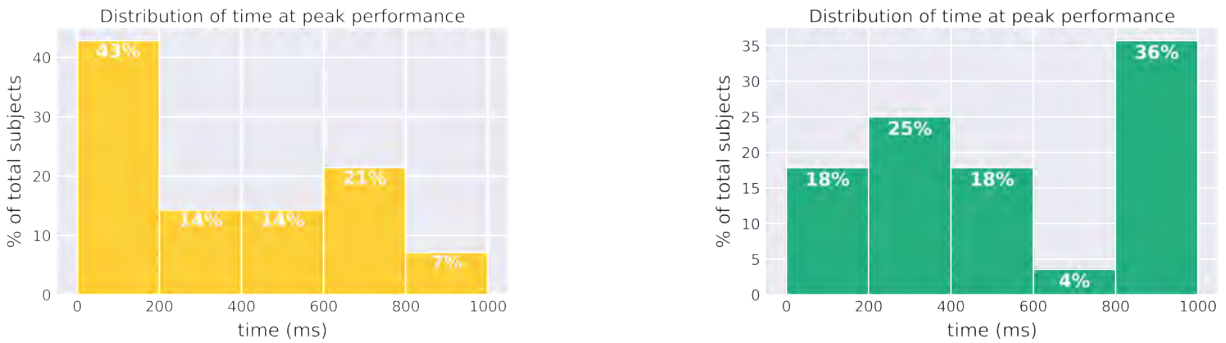
Figure 5.3: Decoding using SVM across time using a 50ms sliding time windows for train and testing across 1000ms trial. (a) semantically related and (b) phonologically related interference conditions. The figures show the distribution of time at which the peak test accuracy occurs for (c) semantic and (d) phonological interference conditions. (svm=support vector machine)

Notably, our results show contextual interference is more decodable than interference due to the phase reversal shown in Chapter 4. To ensure this difference is not due to changes in the cross-validation scheme, we repeated analysis from Section 4.3 using the three-fold, 2-3 repetition averaging scheme (Section 5.2) and did not find significant improvements in phase reversal classification.

5.3.2 Temporal dynamics elicited by contextual similarity

To address part two of the chapter aim–investigating whether the neural signatures of semantic and phonological conflict exhibit distinct time courses, we implemented a 50ms sliding-window
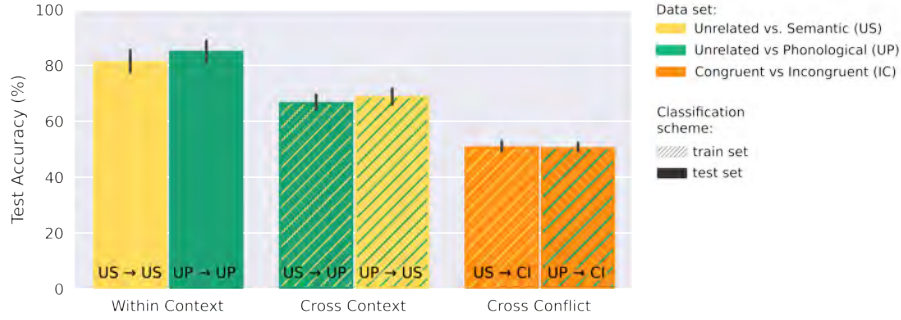
Figure 5.4: Cross-conflict classification accuracies for three data sets: Unrelated vs. Semantic (US), Unrelated vs Phonological (UP), and Congruent vs Incongruent (IC). The classification scheme is indicated by the color of the bars, with the train set in solid colors and the test set in textured patterns. Within-context classification (training and testing on the same conflict type) yields the highest accuracies, followed by cross-context (training on one conflict type and testing on another) and cross-conflict (training on contextual conflict and testing on Stroop-like conflict, or vice versa) classifications.

classification scheme using SVM. The results in Figure 5.3 indicate that the majority of participants showed early classification of semantically-related interference, with control most decodable prior to 600ms after picture presentation (Figure 5.3a). In comparison, the phonologically related condition) showed a later peak in decoding accuracies distributed up to 1000ms after picture presentation (Figure 5.3b). This finding suggests some temporal dissociation between semantic and phonological processing stages. Additionally, the variability in peak time points indicates that participants implement control on varying timelines.

Based on these results, the classification of control recruited under semantic and phonological similarity conditions are both independently viable. Together with the findings in Chapter 4, our preliminary work supports examining cross-condition classification to inform our understanding of generalizable processes in cognitive control within high-similarity conditions.

5.3.3 Generalizability of contextual interference

To address the third and fourth aims, examining the consistency of conflict representations across the generalizability of these representations across different sources of conflict (semantic vs phonological, contextual similarity vs Stroop-like), we completed a cross-contextual and cross-conflict interference classification. In Figure 5.4, we find that the results of within-context classification achieve the highest accuracies, indicating that the neural representations of conflict are consistent across speakers within each conflict type. The lower but above-chance accuracies for

cross-context and Cross-conflict classifications suggest partial generalizability of conflict representations across different sources of interference, with greater similarity between semantic and phonological conflict than between contextual similarity and Stroop-like conflict.

Overall, these results provide valuable insights into the neural dynamics of conflict detection and resolution during language production, supporting the aims of Chapter 5. The results highlight the differential time courses of semantic and phonological conflict, the consistency of conflict representations across speakers, and the partial overlap between different sources of conflict, informing theories of language production and cognitive control.
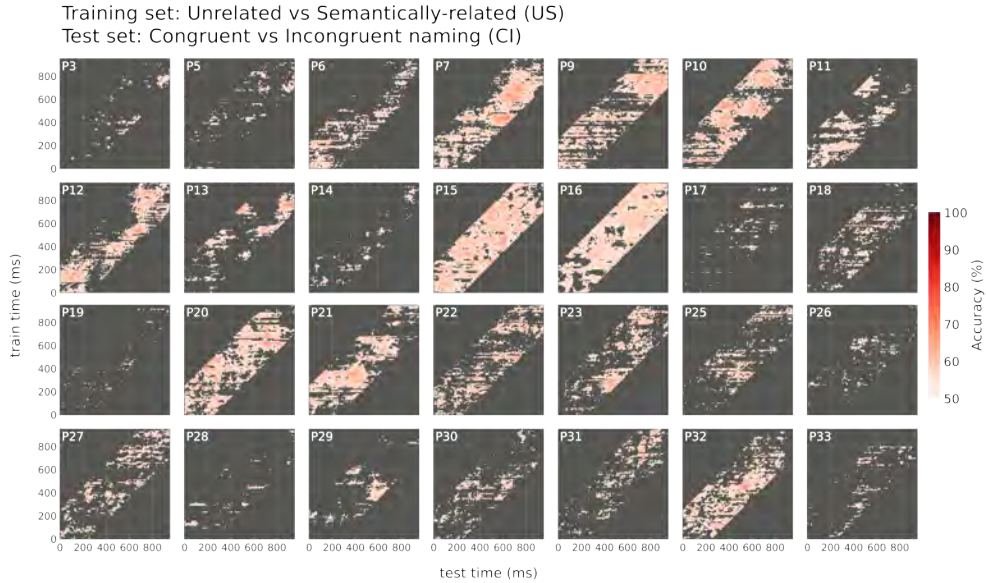
Figure 5.5: Temporal Generalization per participant, across cognitive control domain: Classifiers trained to decode semantically-related vs unrelated picture-word trials. Evaluated on classification accuracy of congruent vs. incongruent trial labels.

5.4   Discussion

In this chapter, we investigated the spatiotemporal dynamics of interference due to contextual similarity during language production using time-resolved decoding of EEG signals. Our results demonstrate that low and high conflict states induced by semantic and phonological similarity can be reliably decoded from neural activity patterns. The successful decoding of these states suggests that the brain engages distinct neural mechanisms to resolve interference arising from the co-activation of semantically and phonologically related word representations.

The time-resolved decoding analysis revealed different temporal profiles for semantic and phono-
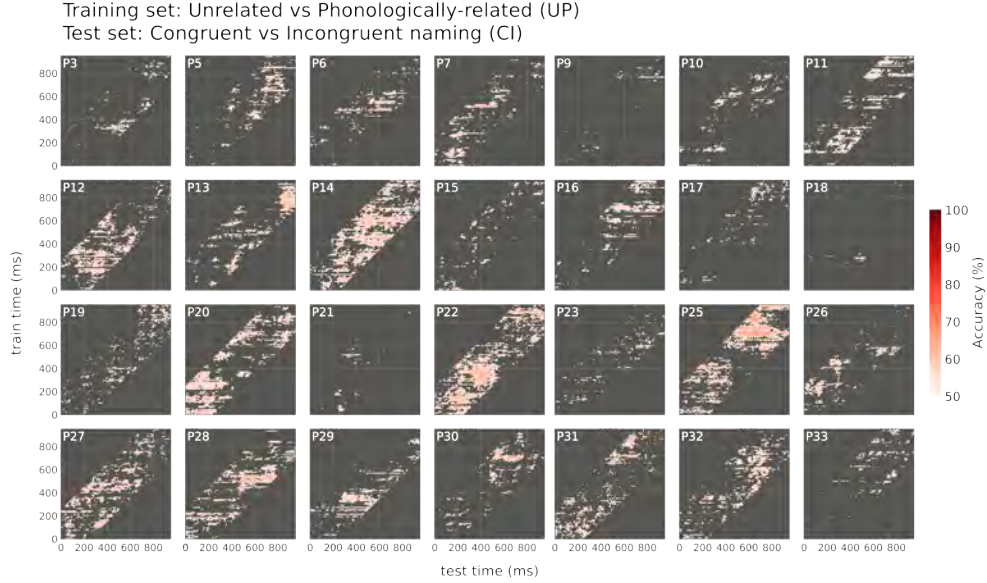
Figure 5.6: Temporal Generalization per participant, across cognitive control domain: Classifiers trained to decode phonologically-related vs unrelated picture-word trials. Evaluated on classification accuracy of congruent vs. incongruent trial labels.

logical interference, with semantic interference being most decodable earlier in the trial compared to phonological interference. This finding supports the idea of a temporal dissociation between semantic and phonological processing stages, consistent with models proposing a sequential flow of information from semantic to lexical to phonological levels. However, the variability in peak decoding times across participants highlights the importance of considering individual differences in the temporal dynamics of language production processes.

Furthermore, our cross-context and cross-conflict decoding analyses shed light on the generalizability of interference representations across different sources of contextual similarity. The higher decoding accuracies for within-context classification compared to cross-context and cross-conflict classification suggest that while there are shared neural patterns underlying interference resolution, there are also distinct representations specific to each type of contextual similarity. These findings contribute to our understanding of the common and unique neural mechanisms involved in resolving interference during language production. Future research should further investigate the factors influencing the generalizability of these mechanisms across individuals and linguistic contexts, as well as their potential implications for models of language production and cognitive control.

Chapter 6

Aim 3

6.1   Introduction

In Chapters 4 and 5, we successfully applied machine learning tools to decode low- and high-conflict states from EEG recordings across varying language-related conditions, including conflict induced by incongruent stimuli and target pairs as well as effects of contextual similarity at different stages of production. In this chapter, we expand our investigation both theoretically and methodologically. Theoretically, we ask whether high and low-conflict states and the consequent control mechanisms share a neural signature between linguistic and non-linguistic tasks. Methodologically, we employ fMRI, which allows us to directly compare the neural correlates of control across domains. The question of whether cognitive control is domain-general or domain-specific remains controversial. Even using the same methodology, such as the congruency sequence effect (CSE; Gratton et al., 1992), some studies have not found the transfer of control between domains (e.g., Freund & Nozari, 2018; Verguts & Notebaert, 2008, 2009), while others have (e.g., Hsu & Novick, 2016). We address this issue by using a paradigm based on Freund and Nozari (2018), which combines two established tasks to elicit control in different domains. Specifically, we examine previously collected fMRI data of participants completing a picture-word interference (PWI) task, which elicits control in the language domain, interleaved with a Simon task, requiring control in the visual-motor domain. By leveraging fMRI's high spatial resolution, we are able to ask where control occurs in the brain. Using machine learning classifiers, we investigate whether the neural correlates of control are shared between a language and non-language task. We ask three questions:

1. Can we successfully classify low vs. high conflict states in PWI using fMRI data?

2. Can we successfully classify low vs. high conflict states in the Simon task?

3. Can we cross-classify low vs. high conflict states between the two tasks?

6.2 Methods

6.2.1 Experimental Paradigm

During the experimental task, participants completed two interleaved cognitive tasks: (1) picture-word interference (PWI), a linguistic task (Schriefers et al., 1990), and (2) Simon task, a non-linguistic visuospatial task (Simon & Rudell, 1967). Each task included congruent and incongruent conditions which required the inhibition of a prepotent response (Hirschfeld et al., 2008; Schriefers et al., 1990). During PWI trials, participants were presented with a line drawing (picture) and a superimposed word, and tasked with naming the picture (target). In congruent trials, the superimposed word corresponded with the picture name, while in incongruent trials, the superimposed word did not match the picture name and acted as a distractor.

The Simon task has been well-validated to produce behavioral effects of increased conflict, such as increases in reaction time, error rate, and response duration (Simon & Rudell, 1967; Hommel, 2011). It has been utilized to examine the neural correlates of cognitive control processes, particularly response selection and conflict resolution. In fMRI studies of the Simon task, the dorsolateral prefrontal cortex (DLPFC) has been implicated in cognitive control, and the anterior cingulate cortex (ACC) has been suggested to monitor for conflict and an increased need for control (Egner & Hirsch, 2005; Fan et al., 2003; Liu et al., 2004). Based on previous studies, the Simon task is an appropriate experimental paradigm for examining cognitive control in the brain within the motor and visuospatial domain.

We aim to investigate the domain-generality or specificity of cognitive control between language production and visual-spatial tasks using multivariate pattern analysis (MVPA). This approach allows for localization of regional activation and identification of distributed patterns of activity within and across tasks. We will use beta values estimated for each voxel at each trial as input, and perform both within- and across-domain classification at both the whole brain and searchlight level (Kriegeskorte et al., 2006).
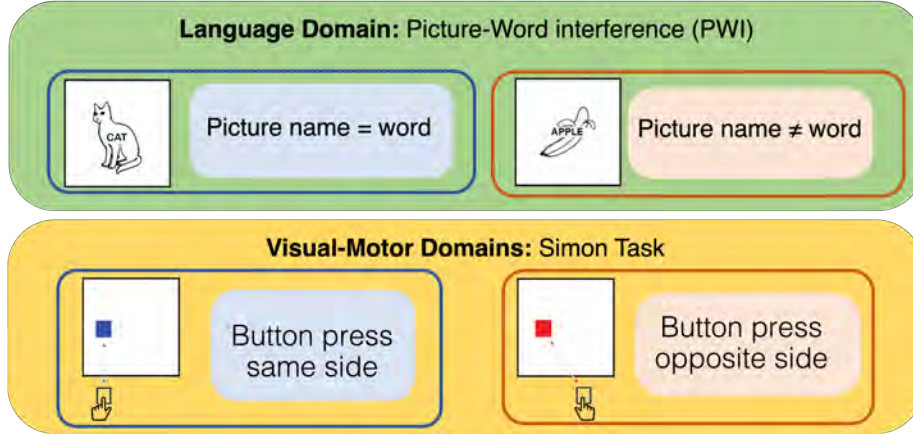
Figure 6.1: Experimental paradigm per cognitive domain: Example of congruent (left) and incongruent (right) trials per Picture-Word Interference (top) and Simon task (bottom).

6.2.2   fMRI Preprocessing and Methods

The acquisition of functional and structural MRI scans for this study was completed by a member of the Nozari Lab. For preprocessing, the fMRI data in each experimental block was z-scored along voxels across TRs. The resulting fMRI volumes are masked with the brain mask, which is created using the bbregister transform for alignment and Freesurfer mask using pycortex. The functional data was motion and slice-timing corrected and spatially transformed to a standard MNI 152 brain template, the resulting output is an fMRI array with dimensions (number of masked voxels × number of TRs from all blocks), which was used in calculating beta coefficients to perform the subsequent classification experiments. Note that this array is transposed for calculating beta coefficients from OLS.

Preliminarily, we completed a univariate analysis, a common approach for fMRI studies, to examine the neural correlates of control of each cognitive domain. The univariate test compares the blood oxygenation level-dependent (BOLD) signal between the specified conditions on a voxel-by-voxel basis. The BOLD signal indirectly measures neural activity by reflecting the changes in blood oxygenation in response to the metabolic demands of a brain region. The preliminary preprocessing of the fMRI data was done collaboratively within the Nozari Lab.

The preprocessed data was then analyzed using a general linear model (GLM) to model the changes in blood oxygenation as a linear combination of various conditions, including experimental factors, noise, and motion. The GLM results in a set of beta coefficients which estimate the contributions from the task condition at each voxel. The beta values were calculated for experiments described below:

Task-level beta values:

(format:Experiment Type $_{\text{condition}}$)

- Incongruent$_{PWI}$

- Congruent$_{PWI}$

- Incongruent$_{Simon}$

- Congruent$_{Simon}$

- (Incongruent + Congruent)$_{PWI}$

- (Incongruent + Congruent)$_{Simon}$

- Incongruent$_{PWI}$ + Incongruent$_{Simon}$

- Congruent$_{PWI}$ + Congruent$_{Simon}$

To determine whether the BOLD signal differs significantly between experimental conditions, a t-test is performed on the beta coefficients. The t-test compares the means of the beta coefficients between two conditions and assesses whether the difference is statistically significant. The resulting t-statistic is then converted to a p-value, which represents the probability of observing the difference in means by chance, under the null hypothesis that there is no true difference between the conditions. Performing multiple t-tests across thousands of voxels increases the risk of false positives, i.e., detecting significant differences that are merely due to chance. To control for this multiple comparisons problem, the false discovery rate (FDR) correction is often applied. The FDR correction adjusts the p-values to maintain a desired false discovery rate, which is the expected proportion of false positives among all significant results.

We completed a univariate analysis to examine the neural correlates of control within each cognitive domain. We calculated the beta coefficients associated with each task and performed group-wise t-tests to compare the BOLD signal between incongruent and congruent trials, both within and across tasks. The following contrasts were analyzed:

Task-level contrasts:

- PWI: (Incongruent$_{PWI}$ − Congruent$_{PWI}$)

- Simon: (Incongruent$_{Simon}$ − Congruent$_{Simon}$)

Experiment-wide contrasts:

- Combined tasks:

$$(\text{Incongruent}_{PWI} + \text{Incongruent}_{Simon}) - (\text{Congruent}_{PWI} + \text{Congruent}_{Simon})$$

- Between tasks: PWI – Simon:

$$(\text{Incongruent}_{PWI} + \text{Congruent}_{PWI}) - (\text{Incongruent}_{Simon} + \text{Congruent}_{Simon})$$

The task-level contrasts aimed to identify the neural correlates of conflict associated with incongruent trials within each cognitive domain, while the experiment-wide contrasts examined the overall effect of conflict across domains and the differences in neural activity between the PWI and Simon tasks, independent of trial congruency. The resulting t-maps were corrected for multiple comparisons using FDR to ensure that the significant activations were not merely due to chance.

Decoding Framework

For the language production classifiers, we took as input the beta values for single trials during the PWI task. The output consisted of two labels: congruent and incongruent. In the within-domain condition, we trained and tested on the PWI task data in cross-validation. In the across-domain condition, we trained on all the PWI task data and tested on all the Simon task data. We performed these analyses at the whole brain and searchlight levels.

Visuo-motor Classifiers

In these analyses, we took as input the beta values for single trials during the Simon task. The output consisted of two labels: congruent and incongruent. In the within-domain condition, we trained and tested on the Simon task data in cross-validation. If necessary, we averaged the trials to improve the signal-to-noise ratio. In the across-domain condition, we trained on all the Simon task data and tested on all the PWI task data. We performed these analyses at the whole brain and searchlight levels.

Preparation of design matrices

For each block, timing information was extracted from scanner logs, including stimuli presentation times along with their labels and TR frame times. Unique events per subject (i.e., trials with the same target and distractor pair stimuli) were defined, resulting in 164 unique events: 160

PWI and 4 Simon target and distractor pairs. It's important to note that the same set of unique events was used across all blocks, and this process did not take into account 1-back or 2-back information. A matrix of TR times × unique events was created and convolved with a (two gamma) hemodynamic response filter. The matrices from each block were then concatenated, resulting in a design matrix per subject with dimensions (number of TRs from all blocks × number of unique events).

Calculation of Beta coefficients

Beta coefficients were estimated using the ordinary least squares (OLS) method:

$$\hat{\beta} = (X^T X)^{-1} X^T y,$$

where X was the design matrix (number of TRs × number of unique events), and y was the transposed functional data (number of TRs × number of voxels). The resulting coefficients had dimensions (number of unique events × number of voxels). The beta coefficients were then separated by task into PWI betas and Simon betas. The task was to decode congruent vs. incongruent labels from beta coefficients associated with unique PWI events. A nested 5-fold cross-validation scheme was utilized to model SVC and linear ridge classifiers independently. The mean classification accuracy per subject was calculated. The resulting output was the classification accuracy of decoding PWI congruent vs. incongruent trials.

Least Squares All (LSA) Calculation of Beta Coefficients

Least Squares All (LSA) is a method for estimating the beta coefficients within the general linear model (GLM) framework that allows for variability in the response across trials of the same type. In the LSA approach, each trial is modeled as a separate regressor in the GLM, enabling the estimation of trial-specific responses. This is in contrast to the traditional GLM approach, where trials of the same type are typically modeled using a single regressor, assuming a consistent response across those trials. By allowing for trial-specific estimates, LSA can capture potential variability in the BOLD response that may be present due to factors such as attentional fluctuations, learning effects, or other trial-specific influences. The LSA approach aims to minimize the squared error across all regressors, resulting in a more fine-grained representation of the neural activity associated with each trial.

Searchlight

A searchlight classifier in the context of fMRI is a multivariate pattern analysis technique that examines the information content of local cubic regions throughout the brain the small "searchlight" through the cortical layer, upon which a classifier is trained and tested only the voxels within the cube. This process is repeated on all possible cortical locations to create a map of classification accuracies that reflects the amount of task-relevant information present in each local region.

In the current study, we performed searchlight analyses to investigate the spatial distribution of information related to cognitive control in the PWI and Simon tasks. Specifically, we trained and tested classifiers to distinguish between congruent and incongruent trials within each task, using the beta coefficients from the voxels within each searchlight as features. By examining the classification accuracy maps obtained from these analyses, we aimed to identify brain regions that contain task-relevant information and potentially shed light on the neural substrates of cognitive control in linguistic and non-linguistic domains.

- Searchlight size: $4 \times 4 \times 4$ voxels

- Classification algorithm: Support Vector Machine (SVM) for both within- and cross-task classification

- Evaluation metric: Group t-test of test accuracy across subjects

ROI decoding of left IFG

The left inferior frontal gyrus (IFG) is a brain region located in the frontal lobe, which has been implicated in various aspects of language processing, including speech production, syntax, and phonological processing. In the context of fMRI studies on cognitive control, the IFG has been shown to play a role in the resolution of competition between co-activated representations, such as in the case of word selection during language production tasks. In this study, we performed region of interest (ROI) analyses focusing on the left IFG to investigate its involvement in cognitive control processes during the PWI. Specifically, we trained and tested classifiers to distinguish between congruent and incongruent trials from beta coefficients associated with left IFG ROI voxels. By examining the classification accuracies obtained from these analyzes, our objective was to assess the extent to which the left IFG contains information relevant to the resolution of conflict in the linguistic domain.

- ROI definition: Left inferior frontal gyrus

- Classification algorithm: Support Vector Machine (SVM)

- Input data: Beta coefficients from voxels within the left IFG ROI

- Evaluation metric: Classification accuracy of congruent vs. incongruent trials

6.3   Results

Task-level univariate test of congruency effect

Figure 6.2 shows the results of the task-level univariate test of congruency effect. For the PWI task (Figure 6.2a), significant clusters of activation were observed in the left inferior frontal gyrus and the anterior cingulate cortex. In contrast, the Simon task (Figure 6.2b) did not show any significant clusters of activation when comparing incongruent and congruent trials.

Experiment-wide univariate tests

The experiment-wide univariate tests (Figure 3) revealed significant clusters of activation in the left inferior frontal gyrus, anterior cingulate cortex, and the right dorsolateral prefrontal cortex when comparing incongruent and congruent trials across both tasks combined. Additionally, significant clusters were observed in the left inferior frontal gyrus and the anterior cingulate cortex when contrasting the PWI and Simon tasks.

Whole-brain searchlight: group-wise classification of trial congruency

Figure 6.4 shows the results of the whole-brain searchlight analysis for group-wise classification of trial congruency. The classification accuracy map (Figure 4a) shows clusters of high accuracy in the left inferior frontal gyrus and the right dorsolateral prefrontal cortex for the within-task classification of PWI trials. However, the FDR-corrected t-score map (Figure Figure 6.4b) did not show any significant clusters for any of the other classification tasks.

Reverse cumulative distribution of voxels with respect to accuracy

Figure 6.5 demonstrates the reverse cumulative distribution of voxels with respect to accuracy on the Simon classification task and the transfer classification from PWI to Simon. Although no significant cross-subject probability map was obtained when combining the results, each subject showed a trend of having a sizeable number of voxels with accuracy above 0.58. However, the number of voxels with an accuracy higher than 0.55 for both tasks was very small and comparable to a random baseline.

cluster threshold: 123 voxels
p < 0.005
α < 0.05

$Incongruent_{PWI} - Congruent_{PWI}$

(a)

cluster threshold: 10 voxels
p < 0.05
α > 0.1

$Incongruent_{Simon} - Congruent_{Simon}$

(b)
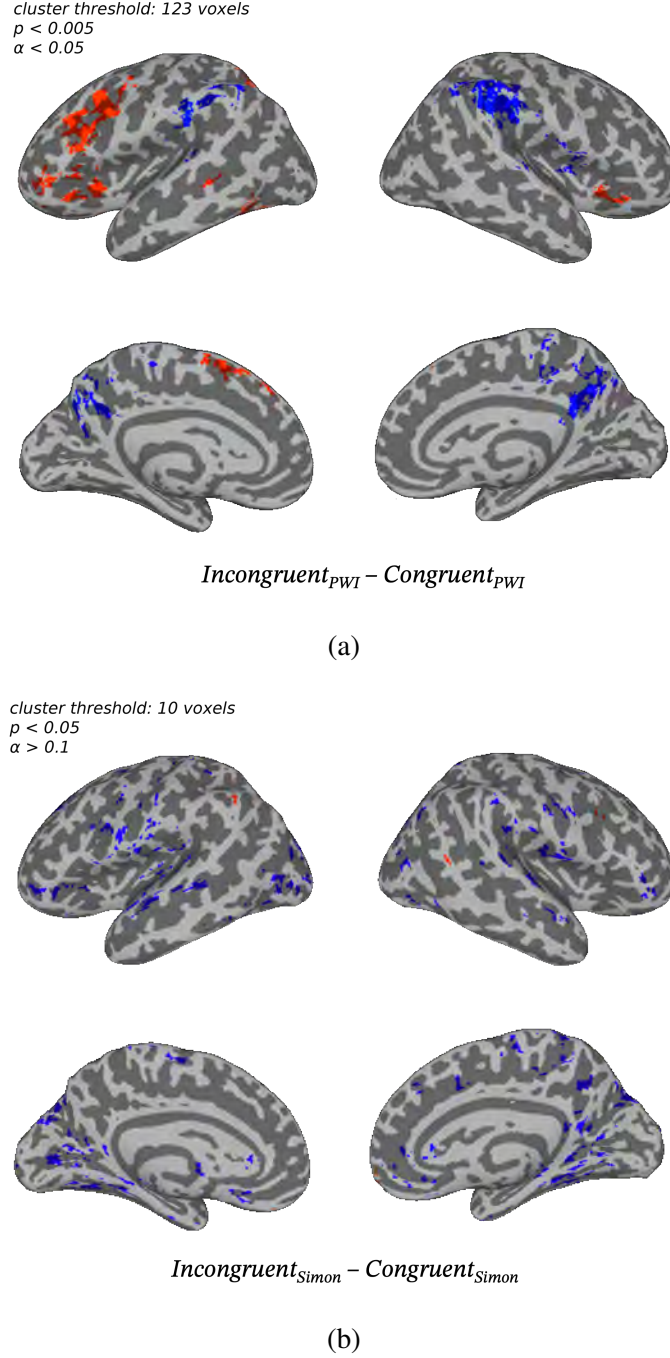
Figure 6.2: Task-level univariate test of congruency effect. (a) t-test results from $(Incongruent_{PWI} - Congruent_{PWI})$ (b) t-test results from $(Incongruent_{Simon} - Congruent_{Simon}$

Searchlight classification accuracy within and across domains for example participants

Figure 6.6 presents the searchlight classification accuracy within and across domains for four example participants. Notable variation was observed in the subject-level classification maps,

cluster threshold: 253 voxels
$p < 0.005$
$\alpha < 0.05$

L   R

$[Incongruent + Congruent]_{Simon} - [Incongruent + Congruent]_{PWI}$

cluster threshold: 142 voxels
$p < 0.005$
$\alpha < 0.05$

L   R

$(Incongruent_{PWI} + Incongruent_{Simon}) - (Congruent_{Simon} + Congruent_{PWI})$

Figure 6.3: Experiment-wide univariate tests

particularly in the first two columns, which do not show a clear pattern across participants in regions that are classifiable with our models.

6.4 Discussion

In this chapter, we investigated the neural correlates of cognitive control in linguistic and non-linguistic domains using fMRI data from participants completing a picture-word interference (PWI) task interleaved with a Simon task. We performed univariate analyses to identify brain regions showing significant differences between incongruent and congruent trials within each task and across both tasks combined. Additionally, we employed multivariate pattern analysis (MVPA) techniques, including whole-brain searchlight and region of interest (ROI) analyses, to examine the classification accuracy of congruency conditions within and across tasks. Our results showed that significant classification was possible only for the within-task PWI classification models, while the Simon task and cross-task classifications did not yield significant results.

One potential cause for the lack of significant results in the Simon task and cross-task classifications is the presence of sequential congruency effects, where the Simon effect is modulated by the compatibility of the prev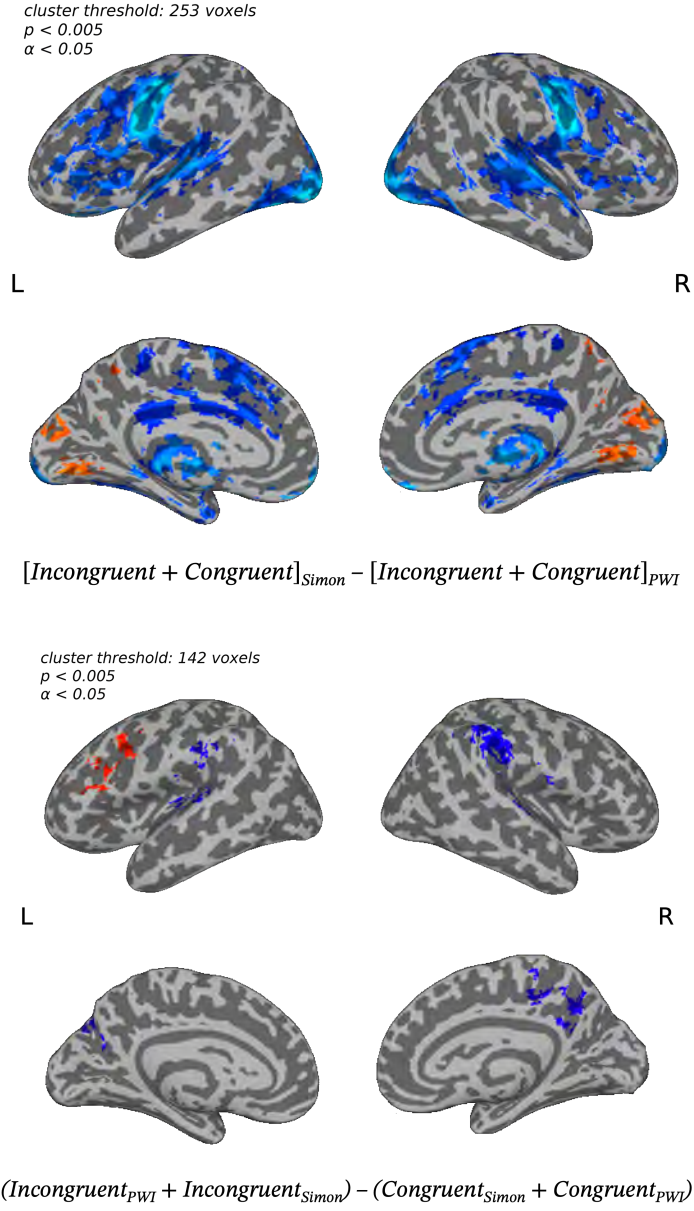ious trial. Previous studies have shown that the Simon effect is typically larger after a congruent trial than after an incongruent trial (Egner, 2007; Stürmer et al., 2002), suggesting that the cognitive control system dynamically adapts to the level of conflict experienced on the previous trial. Kerns et al. (2006) found that the anterior cingulate cortex (ACC) showed higher activity on incongruent trials preceded by incongruent trials (iI trials) compared to incongruent trials preceded by congruent trials (cI trials), and this ACC activity was associated with increased prefrontal cortex (PFC) activity on the next trial. Similarly, Horga et al. (2011) reported that the dorsal premotor cortex (dPMC) showed higher activity on congruent trials preceded by incongruent trials (iC trials) compared to congruent trials preceded by congruent trials (cC trials). These findings highlight the importance of considering sequential congruency effects in the analysis of fMRI data from tasks involving cognitive control.

Given that our fMRI experiment, the interleaved design of the PWI and Simon tasks, with congruent and incongruent trials in the 1-back and 2-back positions, may have introduced additional

Classification	Train Set	Test Set	t-value	p-value
Within-task	PWI	PWI	4.519	0.0002*
	Simon	Simon	-1.745	0.095
Cross-task	PWI	Simon	0.505	0.6184
	Simon	PWI	1.106	0.2806

Table 6.1: Classification accuracy in the Left inferior frontal gyrus for four classification experiments. The results indicate only significant accuracy of the IFG classification in the classification within seconds of congruent and incongruent trials.
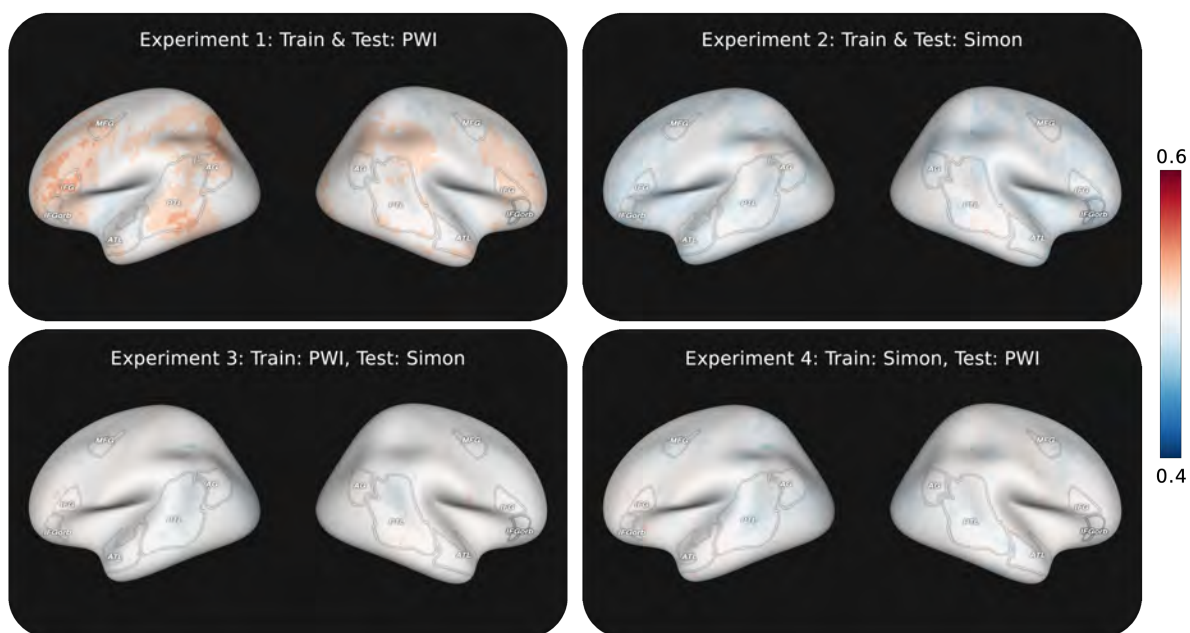
(a)
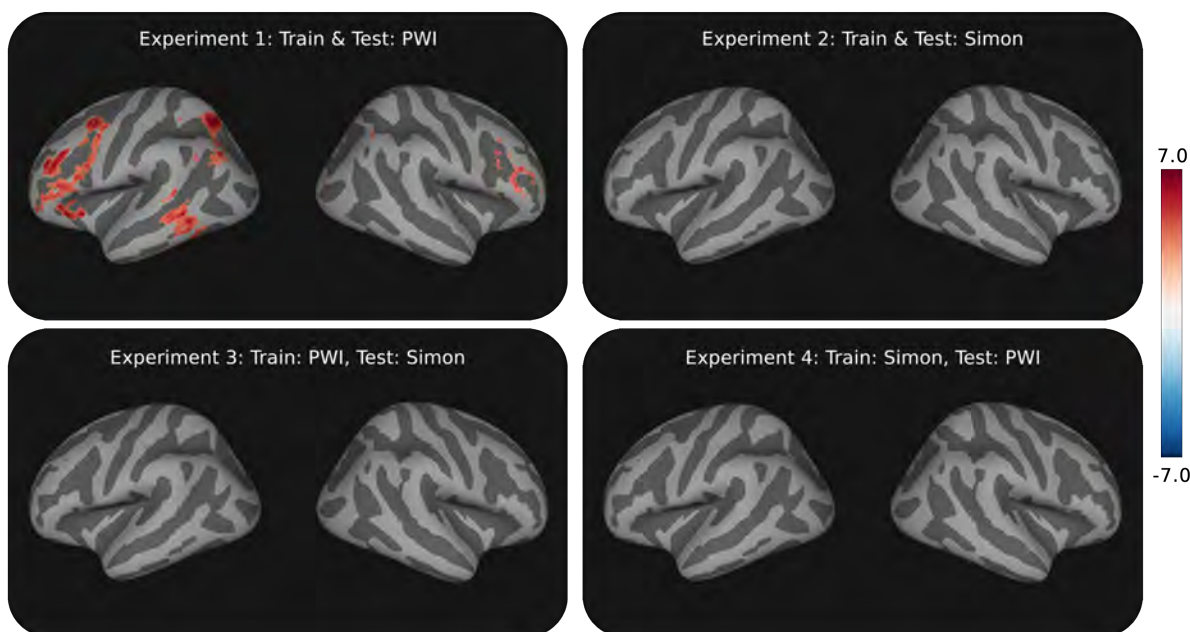
(b)

Figure 6.4: Whole-brain searchlight: group-wise classification of trial congruency. (a) Classification accuracy and (b) FDR-corrected t-test($\alpha = 0.05$). In Experiment 1, a within-task experiment, training and testing on PWI trials showed clusters of significance in prefrontal, and parietal regions for distinguishing congruent vs incongruent trials. Experiment 2, 3, and 4 did not show significant clusters.
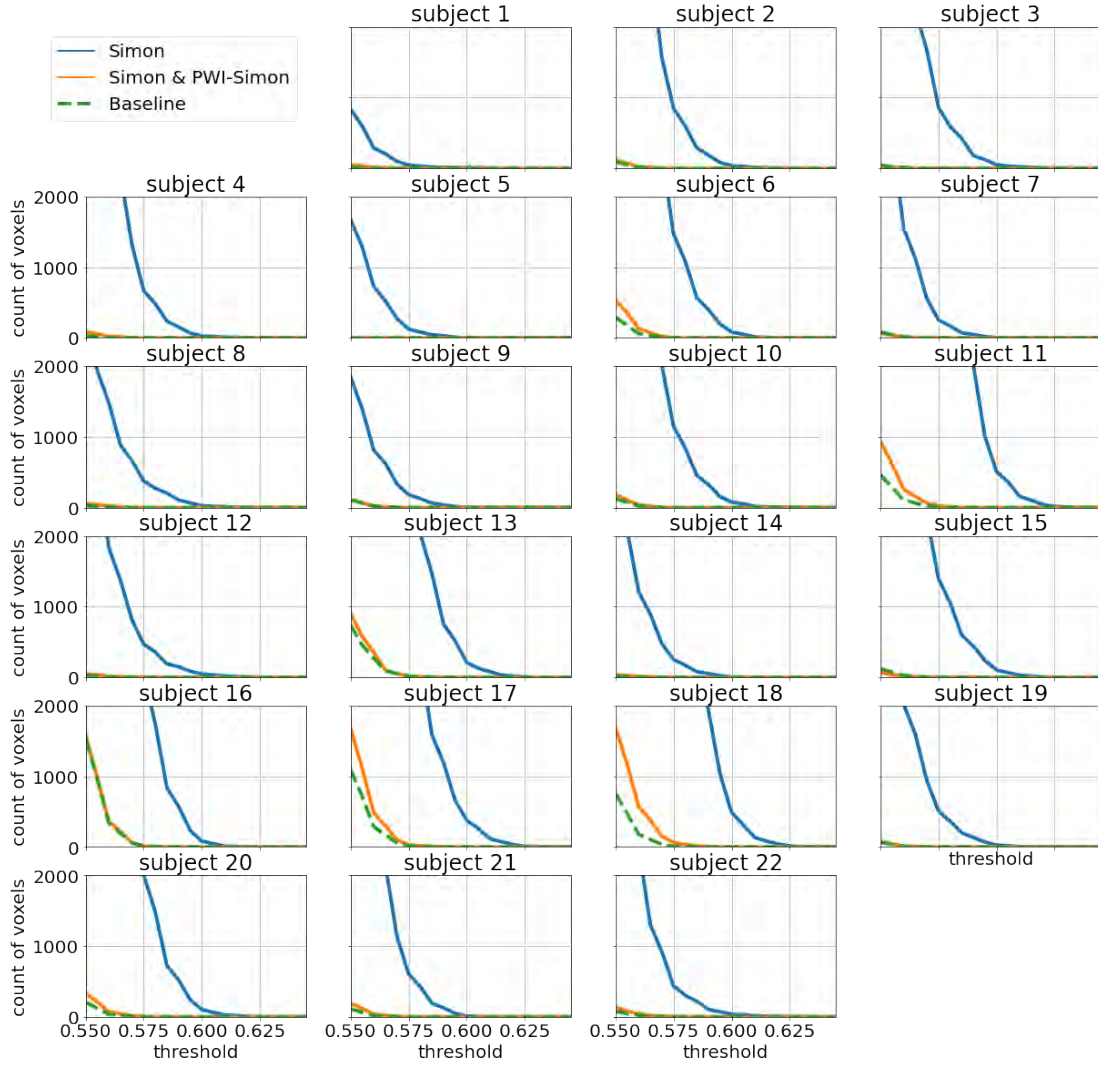
Figure 6.5: Reverse cumulative distribution of voxels with respect to accuracy on the Simon classification task and the transfer classification from PWI to Simon. For each subject and each accuracy value (e.g., 0.6) the number of voxels with Simon classification accuracy at least that large (e.g., voxels with accuracy larger than 0.6) was obtained. The number of voxels with accuracy at least that large for both the Simon classification task and the transfer classification from PWI to Simon was also obtained. We see that even though we were not able to obtain a significant cross-subject probability map when combining the results, each subject has a trend of having a sizeable number of voxels with accuracy above 0.58. The number of voxels which have an accuracy higher than 0.55 for both the tasks is very small, and is comparable with a random baseline in which we multiplied the proportion of voxels from Simon with the average proportion of the transfer task on the brain.

complexity in the neural responses to conflict.

Another factor that may have contributed to the lack of significant results is the subject-level
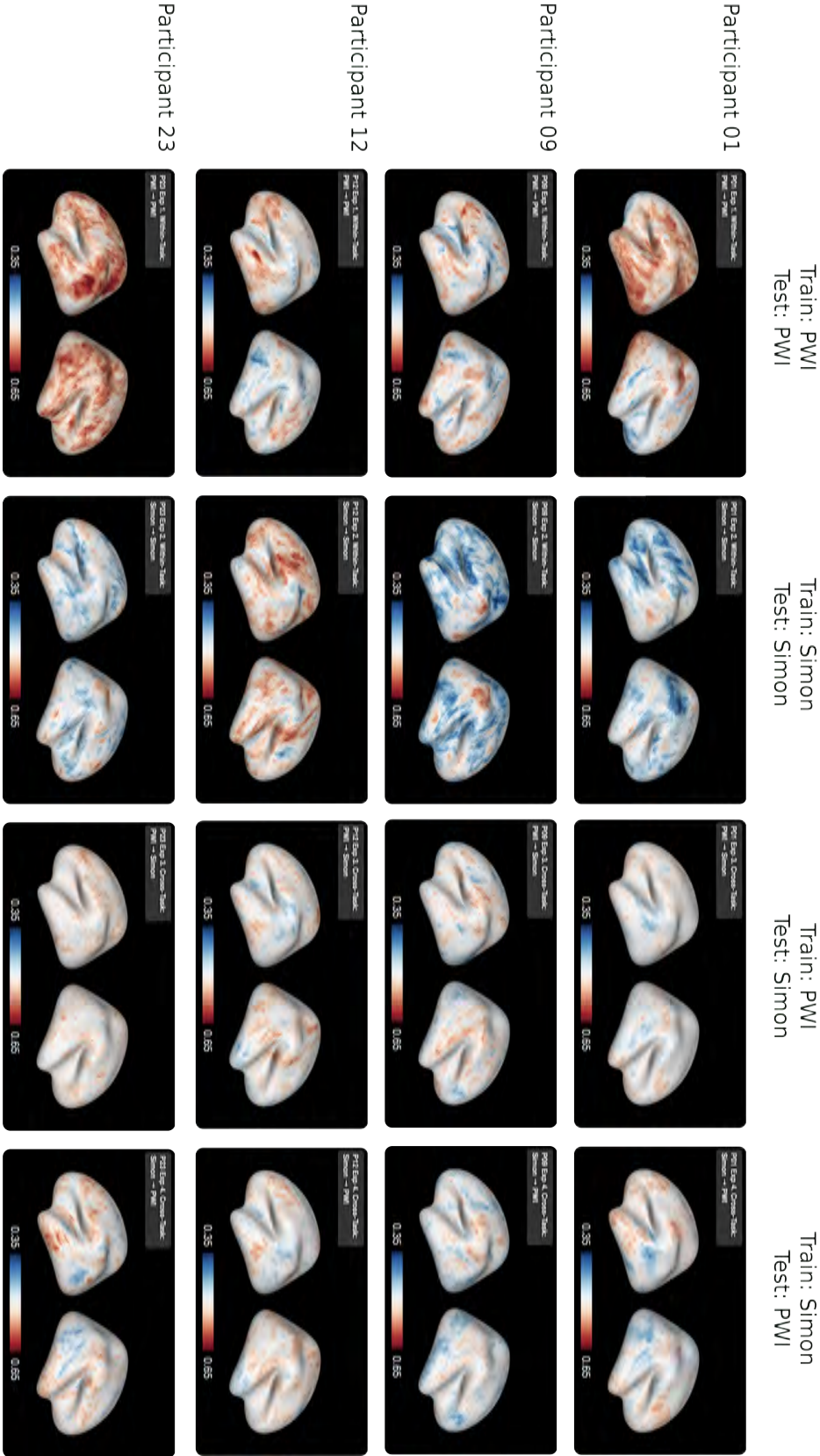
Figure 6.6: Searchlight classification accuracy within and across domains for four example participants (P01, P09, P12, P23). Columns are organized by the task type that was utilized for model training and testing. (left to right) Columns: (1) within-task classification of congruent vs incongruent PWI trials; (2) within task classification of congruent vs incongruent Simon trials; (3) across-task classification of congruent vs incongruent trials trained on PWI task, Tested on Simon task trials; (4)across-task classification of congruent vs incongruent trials trained on Simon task, Tested on PWI task trials

variations in the brain regions involved in cognitive control. As shown in Figure 6, there was notable variation in the searchlight classification accuracy maps across participants, particularly in the within-task classifications for the PWI and Simon tasks. This variability suggests that individuals may employ different neural strategies or engage different brain regions when resolving conflict in the visuo-motor domain. These individual differences in cognitive control strategies may have reduced the sensitivity of group-level analyses to detect consistent patterns of activation or classification accuracy across participants.

Furthermore, our univariate analysis focused primarily on identifying brain regions showing increased activation during incongruent trials compared to congruent trials. However, it is possible that task-related deactivations may also play a crucial role in cognitive control processes. The Petruo study (2021) on Simon task decoding, found that task set deactivation in specific regions, such as the left orbitofrontal and inferior parietal cortices, was more predictive of performance than task set activation. Future analyses should also leverage this knowledge to investigate gain a more comprehensive understanding if there are unique patterns of deactivation vs activation when examining linguistic and non-linguistic domains.

6.5 Future steps

One important next step in addressing the concerns raised in the Discussion section is to examine the sequences of congruent and incongruent trials in the interleaved PWI and Simon tasks. By considering the congruency of the 1-back and 2-back trials, we can investigate how the congruency effect changes depending on the preceding trial types. This analysis would involve creating new models that take into account the different possible sequences of congruent and incongruent trials across the two tasks. For example, we could define eight different trial types based on the congruency of the current trial, the 1-back trial, and the 2-back trial, as well as the task type (PWI or Simon).

By comparing the neural activity and classification accuracy across these different trial types, we can gain insights into how the congruency effect is modulated by the preceding trial types in both the PWI and Simon tasks. This analysis would help us better understand the dynamic nature of cognitive control processes and how they are influenced by the specific sequence of congruent and incongruent trials across different task domains. By comparing the neural activity and classification accuracy across these different trial types, we can gain insights into how the congruency effect is modulated by the preceding trial types in both the PWI and Simon tasks. This analysis would help us better understand the dynamic nature of cognitive control processes

and how they are influenced by the specific sequence of congruent and incongruent trials across different task domains.

Chapter 7

Conclusion

7.1 Scientific contributions

7.1.1 Supressing prepotent responses

The findings presented in Chapter 4 make significant contributions to our understanding of the neural mechanisms underlying prepotent response suppression in language production. By employing a Stroop-like picture-naming task and applying machine learning techniques to EEG data, we have demonstrated the feasibility of decoding cognitive control states from neural activity patterns. The successful classification of high and low-conflict trials based on EEG signals suggests that there are distinct neural signatures associated with the suppression of prepotent responses. This finding provides empirical support for the involvement of cognitive control processes in language production and highlights the potential of using neural markers to track the engagement of these processes.

Furthermore, the time-resolved decoding analysis in Chapter 4 sheds light on the temporal dynamics of prepotent response suppression. By characterizing the emergence and duration of decodable cognitive control states, we have shown that the neural signatures of control extend beyond the initial stages of lexical selection (Indefrey & Levelt, 2004). This finding challenges the strictly serial view of language production and suggests that cognitive control processes operate in a more continuous manner, influencing multiple stages of word production (Strijkers et al., 2011). The temporal information gained from this analysis can inform models of language production by providing insights into the time course of control mechanisms and their interaction with linguistic processes.

Finally, the examination of the generalizability of cognitive control signatures across participants in Chapter 4 reveals important insights into the commonalities and individual differences in pre-potent response suppression. The above-chance cross-participant decoding accuracy indicates that there are shared neural patterns underlying cognitive control across individuals (Fedorenko et al., 2013). However, the higher performance of within-participant classifiers suggests that individual differences also play a role in the neural implementation of control processes. This finding highlights the importance of considering both group-level and individual-level factors in the study of cognitive control in language production. By characterizing the spectrum of individual variability, we can develop more comprehensive models of language processing that account for the diversity of cognitive control strategies employed by different speakers.

7.1.2 Resolving contextual interference

The research presented in Chapter 5 makes important contributions to our understanding of the neural mechanisms involved in resolving competition during language production. By employing a blocked-cyclic naming paradigm and manipulating the semantic and phonological relatedness of the stimuli, we have demonstrated the different effects of these linguistic relationships on the neural signatures of cognitive control. The successful decoding of high and low conflict states in both semantically and phonologically related conditions suggests that the brain engages control processes to resolve competition arising from the co-activation of related word representations (Schnur et al., 2006). This finding provides neural evidence for the role of cognitive control in mitigating interference during lexical selection.

The time-resolved decoding analysis in Chapter 5 reveals distinct temporal profiles for the resolution of semantic and phonological competition. The earlier peak decoding accuracy for semantically related conditions compared to phonologically related conditions suggests that semantic interference may be resolved at an earlier stage of word production, consistent with models proposing a semantic-to-lexical mapping stage (Levelt et al., 1999). In contrast, the later peak decoding accuracy for phonologically related conditions indicates that phonological interference may persist longer and require control processes at a later stage, potentially during the mapping of lexical representations to phonological forms (Dell, 1986). These findings provide insights into the temporal organization of word production and the stage-specific engagement of cognitive control mechanisms.

Furthermore, comparing of the neural signatures of competitive resolution across different linguistic relationships sheds light on the specificity and generality of cognitive control mechanisms in language production. The successful cross-classification between semantically and phonolog-

ically related conditions suggests that there are shared neural patterns underlying the resolution of competition, regardless of the specific linguistic relationship (Ries et al., 2015). This finding supports the idea of a domain-general cognitive control mechanism that operates across different levels of linguistic processing. However, the lower cross-classification accuracy compared to within-condition decoding suggests that there may also be distinct neural processes specific to each type of linguistic relationship. This finding highlights the importance of considering both the commonalities and differences in the neural implementation of cognitive control across various linguistic contexts.

7.1.3   Interference and resolution across domains

The results from Chapter 6 provide important insights into the neural mechanisms underlying cognitive control in linguistic and non-linguistic domains. While we found significant classification accuracy for within-task PWI trials, the lack of significant results for the Simon task and cross-task classifications highlights the complex nature of cognitive control processes and the potential influence of factors such as sequential congruency effects and individual differences in neural strategies. These findings underscore the need for further research to investigate the dynamic interplay between conflict resolution mechanisms across different task domains and how they are modulated by the specific sequence of congruent and incongruent trials.

7.2   Conclusion

The cumulative findings from Chapters 4 and 5 make significant contributions to our understanding of the neural bases of cognitive control in language production. By combining experimental paradigms that manipulate different aspects of cognitive control with advanced neuroimaging techniques and machine learning approaches, this thesis provides a comprehensive examination of the neural mechanisms underlying prepotent response suppression and competition resolution.

The successful decoding of cognitive control states from neural activity patterns demonstrates the feasibility of using brain signals to track the recruitment of control processes during language production. Furthermore, the identification of specific neural signatures associated with different types of linguistic relationships and control demands can inform the design of targeted interventions for individuals with language impairments, such as those with aphasia or developmental language disorders (Abutalebi & Green, 2007).

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. Journal of Neurolinguistics, 20(3), 242–275.

Altemeier, L. E., Abbott, R. D., & Berninger, V. W. (2008). Executive functions in becoming writing readers and reading writers: Note taking and report writing in third and fifth graders. Developmental Neuropsychology, 33(3), 331–354.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40–79.

Balthazar, M. L. F., Cendes, F., & Damasceno, B. P. (2008). Semantic error patterns on the Boston Naming Test in normal aging, amnestic mild cognitive impairment, and mild Alzheimer's disease: Is there semantic disruption? Neuropsychology, 22(6), 703.

Becker, M., Schubert, T., Strobach, T., Gallinat, J., & Kühn, S. (2011). Simultaneous interpreters vs. professional multilingual controls: Group differences in cognitive control as well as brain structure and function. NeuroImage, 134, 250-260.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. The Quarterly Journal of Experimental Psychology Section A, 58(4), 667–692.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300.

Biegler, K. A., Crowther, J. E., & Martin, R. C. (2008). Consequences of an inhibition deficit for word production and comprehension: Evidence from the semantic blocking paradigm. Cognitive Neuropsychology, 25(4), 493-527.

Birn, R. M. (2007). The role of physiological noise in resting-state functional connectivity.

NeuroImage, 62(2), 864-870.

Bode, S., Feuerriegel, D., Bennett, D., & Alday, P. M. (2019). The Decision Decoding ToolBOX (DDTBOX)–A multivariate pattern analysis toolbox for event-related potentials. Neuroinformatics, 17(1), 27–42.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. Psychological Review, 108(3), 624.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees.

Cahana-Amitay, D., & Albert, M. L. (2015). Neuroscience of aphasia recovery: The concept of neural multifunctionality. Current Neurology and Neuroscience Reports, 15, 1–8.

Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research, 11, 2079–2107.

Cohen, M. X. (2014). Analyzing neural time series data: Theory and practice. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(2), 345-361.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3), 283–321.

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: Behavioral and computational considerations. The Oxford Handbook of Language Production, 88-104.

Duann, J. R., Jung, T. P., Kuo, W. J., Yeh, T. C., Makeig, S., Hsieh, J. C., & Sejnowski, T. J. (2002). Single-trial variability in event-related BOLD signals. NeuroImage, 15(4), 823-835.

Egner, T., & Hirsch, J. (2005a). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nature Neuroscience, 8(12), 1784–1790.

Egner, T., & Hirsch, J. (2005b). The neural correlates and functional integration of cognitive control in a Stroop task. Neuroimage, 24(2), 539–547.

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y.-C., & Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: Leave one subject out. NeuroImage, 50(2), 572–576.

Fahrenfort, J. J., van Driel, J., van Gaal, S., & Olivers, C. N. L. (2017). From ERPs to MVPA using the Amsterdam Decoding and Modeling toolbox (ADAM). Frontiers in Neuroscience, 11, 368.

Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., & Posner, M. I. (2003). Cognitive and brain consequences of conflict. Neuroimage, 18(1), 42-57.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. Proceedings of the National Academy of Sciences, 110(41), 16616–16621.

Fox, M. D., Snyder, A. Z., Zacks, J. M., & Raichle, M. E. (2006). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. Nature Neuroscience, 9(1), 23-25.

Freund, M., & Nozari, N. (2018). Is adaptive control in language production mediated by learning? Cognition, 176, 107–130. Fyshe, A., Talukdar, P., Murphy, B., & Mitchell, T. (2019). A compositional and interpretable semantic space. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 1682–1692.

Garrett, D., Peterson, D. A., Anderson, C. W., & Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2), 141–144.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. Clinical Neurophysiology, 114(9), 1580–1593.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. Journal of Experimental Psychology: General, 121(4), 480.

Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. Psychophysiology, 55(3), e13016.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. Journal of Cognitive Neuroscience, 29(4), 677–697.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data

mining, inference, and prediction.

Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. NeuroImage, 180, 4–18.

Hommel, B. (2011). The Simon effect as tool and heuristic. Acta Psychologica, 136(2), 189–202.

Hsu, N. S., & Novick, J. M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. Psychological Science, 27(4), 572-582.

Im-Bolter, N., Johnson, J., & Pascual-Leone, J. (2006). Processing limitations in children with specific language impairment: The role of executive function. Child Development, 77(6), 1822–1841.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. Cognition, 92(1–2), 101–144.

Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: modelling, inference and optimization. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 354(1387), 1215-1228.

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. Trends in Cognitive Sciences, 18(4), 203–210.

King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2016). Two distinct dynamic modes subtend the detection of unexpected sounds. PLoS One, 11(1), e0144180.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. Proceedings of the National Academy of Sciences, 103(10), 3863–3868.

Lahey, M., & Edwards, J. (1999). Naming errors of children with specific language impairment. Journal of Speech, Language, and Hearing Research, 42(1), 195–205.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22(1), 1–38.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. Journal of Neural Engineering, 4(2), R1–R13.

Luck, S. J. (2014). An introduction to the event-related potential technique.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of Neuroscience Methods, 164(1), 177–190.

McNeil, M., Hula, W., & Sung, J. E. (2011). The role of memory and attention in aphasic language performance. In The handbook of psycholinguistic and cognitive processes (pp. 551–577). Psychology Press.

Melrose, R. J., Campa, O. M., Harwood, D. G., Osato, S., Mandelkern, M. A., & Sultzer, D. L. (2009). The neural correlates of naming and fluency deficits in Alzheimer's disease: An FDG-PET study. International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences, 24(8), 885–893.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience, 24(1), 167–202. Montembeault, M., Chapleau, M., Jarret, J., Boukadi, M., Laforce Jr, R., Wilson, M. A., Rouleau, I., & Brambati, S. M. (2019). Differential language network functional connectivity alterations in Alzheimer's disease and the semantic variant of primary progressive aphasia. Cortex, 117, 284–298.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage, 59(3), 2636-2643.

Munding, D., Dubarry, A. S., & Alario, F. X. (2016). On the cortical dynamics of word production: a review of the MEG evidence. Language, Cognition and Neuroscience, 31(4), 441-462.

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Frontiers in Human Neuroscience, 7, 138.

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. Journal of Memory and Language, 76, 253-272.

Notebaert, W., & Verguts, T. (2008). Cognitive control acts locally. Cognition, 106(2), 1071–1080.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. Cognitive, Affective, & Behavioral Neuroscience, 5(3), 263–281.

Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a "lexical editor" be? Journal of Memory and Language, 60(2), 291–307.

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. Current Directions in Psychological Science, 26(5), 403–410.

Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. Journal of Neurolinguistics, 53, 100875.

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. Frontiers in Psychology, 7, 1097.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. Cognition, 114(2), 227-252.

Piai, V., Roelofs, A., & Schriefers, H. (2012). Distractor strength and selective attention in picture-naming performance. Memory & Cognition, 40(4), 614-627.

Pinet, S., & Nozari, N. (2020). Electrophysiological correlates of monitoring in typing with and without visual feedback. Journal of Cognitive Neuroscience, 32(4), 603-620.

Pinet, S., & Nozari, N. (2023). Different electrophysiological signatures of similarity-induced and Stroop-like interference in language . Journal of Cognitive Neuroscience, 35(8), 1329-1349.

Pinet, S., Dell, G. S., & Alario, F.-X. (2019). Tracking keystroke sequences at the cortical level reveals the dynamics of serial order production. Journal of Cognitive Neuroscience, 31(7), 1030–1043.

Porcaro, C., Balsters, J. H., Mantini, D., Robertson, I. H., & Wenderoth, N. (2019). Removal of muscle artifacts from EEG recordings: A critical comparison of two techniques. Psychophysiology, 56(2), e13315.

Ramus, F., & Szenkovits, G. (2008). What phonological deficit? Quarterly Journal of Experimental Psychology, 61(1), 129–141.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. Psychological Review, 107(3), 460-499.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science, 306(5695), 443–447.

Ridderinkhof, K. R., Van Den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection,

response inhibition, performance monitoring, and reward-based learning. Brain and Cognition, 56(2), 129–140.

Ries, S. K., Pinet, S., Nozari, N. B., & Knight, R. T. (2021). Characterizing multi-word speech production using event-related potentials. Psychophysiology, 58(5), e13788.

Ries, S. K., Schendel, K. L., Herron, T. J., Dronkers, N. F., Baldo, J. V., Turken, & U. (2021). Neural underpinnings of proactive interference in working memory: Evidence from patients with unilateral lesions. Frontiers in Neurology, 12, 607273.

Ries, S. K., Schendel, K., DeMarco, A. T., & Turken, A. U. (2021). The role of ventrolateral prefrontal cortex in language production. Journal of Cognitive Neuroscience, 33(4), 705-724.

Riès, S., Karzmark, C., Navarrete, E., Knight, R., & Dronkers, N. (2015). Specifying the role of the left prefrontal cortex in word selection. Brain and Language, 149, 135–147.

Rifkin, R., & Lippert, R. A. (2007). Notes on regularized least squares. MIT-CSAIL-TR-2007-025. http://hdl.handle.net/1721.1/37318

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage, 23(2), 752-763.

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology, 56(6), e13335.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. Journal of Memory and Language, 54(2), 199–227.

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. Proceedings of the National Academy of Sciences, 106(1), 322–327.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory and Language, 29(1), 86–102.

Simon, J. R., & Rudell, A. P. (1967). Auditory SR compatibility: The effect of an irrelevant cue on information processing. Journal of Applied Psychology, 51(3), 300.

Smith, N. J., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP

framework. Psychophysiology, 52(2), 157–168.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory, 6(2), 174–215.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), 111–133.

Strijkers, K. (2016). A Neural Assembly–Based View on Word Production: The Bilingual Test Case. Language Learning, 66(S2), 92-131.

Strijkers, K., Costa, A., & Thierry, G. (2011). The speed of object naming: A high-density ERP study. Frontiers in Psychology, 2, 371.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18(6), 643–662.

Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications, 37(12), 8659–8666.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7(1), 91.

Varoquaux, G., Raamana, P. R., Engemann, D. A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. NeuroImage, 145, 166–179.

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. Trends in Cognitive Sciences, 13(6), 252–257.