Neural Correlates Of The Production Effect: Insights From Univariate And Multivariate Analyses Of fMRI Data

Lyam Meares Bailey

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

May 2024

DEDICATION PAGE

This thesis is dedicated to Antony William Chase Meares.

A True Gentleman.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The production effect is a simple phenomenon in human memory: words that are read aloud are more likely to be remembered on a subsequent memory test (e.g., a recognition task) compared to words read silently. This effect, along with its proposed underlying mechanisms, has received much attention from cognitive research relying near-exclusively on data from behavioural paradigms. To date, the neural underpinnings of this effect remain largely unexplored. In this dissertation I present a comprehensive functional magnetic resonance imaging (fMRI)-based investigation of the production effect in recognition memory. In three fMRI experiments, participants read a series of words either aloud or silently and (in two experiments) were later tested on their recognition memory for all the studied words. Potential neural correlates of the production effects ought to be evidenced by contrasts of brain activity between aloud and silent reading conditions; therefore, across three empirical chapters, I present evidence of differences in encoding and retrieval processes between these two reading conditions. In Chapter 2 I show that aloud reading (during both encoding and subsequent recognition) elicits preferential activity in brain areas associated with speech production and perception. In Chapter 3, I use representational similarity analysis (RSA) to demonstrate that reading aloud is associated with increased sensitivity to multiple properties of presented words; primarily phonological and articulatory features. In Chapter 4, I use pattern similarity analysis (PSA) to show that recognising aloud versus silent words depends on dissociable roles of neural reactivation (reinstatement of stimulus-specific activity patterns from encoding) and neural transformation (systematic changes to activity patterns). Taken together, these findings are broadly consistent with dominant theoretical accounts of the production effect whereby aloud words benefit from distinctive sensorimotor processing. However, some results are also compatible with alternative accounts which emphasise semantic and attentional components. Moreover, evidence for dissociable processes between aloud and silent reading challenges long-standing assumptions of prior research in this area. Overall this work provides novel insights, at the neural level, into encoding and retrieval processes which contribute to the behavioural production effect.

LIST OF ABBREVIATIONS

BOLD	Blood oxygen level dependent
DLPFC	Dorsolateral prefrontal cortex
DMPFC	Dorsomedial prefrontal cortex
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
i-EEG	Intracranial electroencephalography
LOC	Lateral occipital cortex
MVPA	Multi-voxel pattern analysis
PE	Production effect
Pre-SMA	Pre-supplementary motor area
PSA	Pattern similarity analysis
RSA	Representational similarity analysis
SMA	Supplementary motor area
VTC	Ventral temporal cortex
VVP	Ventral visual pathway
VWFA	Visual word form area

ACKNOWLEDGEMENTS

I would like to thank my academic supervisor, Aaron Newman, for his continued guidance and support throughout my time at Dalhousie. Aaron, you (astonishingly) put up with me throughout a Masters *and* a PhD, which I think took no small amount of patience. You have taught me so much—not just about brain imaging, but about academia, academics, writing, teaching, publishing, programming…suffice it to say that I've learned a lot under your tutelage. Moreover, you have always let me plot my own course, particularly in the pursuit of side projects. Even when my ideas were strange and unusual, or had little-to-nothing to do with brain imaging or language processes, you always gave me the resources I needed and let me take the reins. I do not think I would have flourished in grad school without the freedom you gave me to explore.

To the other members of my supervisory committee—Tracy Taylor-Helmick and Tim Bardouille—thank you for your invaluable support and advice throughout the design, execution, and writing of this thesis. To my external examiner, Brice Kuhl, thank you for taking the time to review my work.

I would also like to thank my closest collaborators, whom over the years I have come to consider my friends: Eve Higby, Jon Fawcett, Heath Matheson, and Tim Bardouille. Each of you has helped me to learn and grow as an academic, and to broaden my horizons beyond the narrow scope of my PhD—from Bilingualism to Bayesian statistics and MEG to Open Science. Not only that, but the occasional get-togethers for sushi, beer, or coffee (even when they were over Zoom) were always a welcome break for a grad student who spent too much time working and not enough time socialising.

To my parents and sisters—Mum, Trevor, Ceri, and Jenna—thank you for always being there for a chat on the phone, for picking me up from train stations and airports, for sending unsolicited pork scratchings and scampi fries (and let's not forget the very much solicited financial support), and for tolerating my endless refrain of "I'm too busy this week / month / summer / Birthday / Christmas" for the past 8 years. I love you all, and thank you.

To my wife, Laura. You have always loved and supported me no matter what—through the stress, the anxiety, the moments of despair. I certainly would not be here without your unwavering kindness and support. You are all at once my rock, my friend, my confidant, and my home. For everything—absolutely everything—thank you. Here's to more video games, more "evening tea", more congratulatory and/or commiseratory sushi, and more shenanigans.

CHAPTER 1 GENERAL INTRODUCTION

1.1. Overview

The goal of this dissertation is to characterise neural correlates of the *production effect* as revealed by functional magnetic resonance imaging (fMRI). The production effect is a behaviourally-defined empirical phenomenon whereby words read aloud are more readily remembered than words read silently (MacLeod et al., 2010). With some exceptions, this effect is typically studied in the context of recognition memory; as such, I will focus on the production effect in recognition specifically. While the production effect has received much attention from a cognitive-behavioural perspective, its neural basis remains largely unexplored. Research in this area is valuable because the production effect speaks to broader principles of human memory; that is, how interactions with other cognitive domains, such as language, can modulate encoding and subsequent remembering. Therefore, understanding this behavioural effect, as well as its neural substrates, will improve our broader understanding of encoding and retrieval in the context of a complex cognitive system in which multiple processes (memory, language, motor control, etc.) interact, giving rise to our internal cognitive states and behaviour.

In this dissertation I will present data from three fMRI experiments which entailed adaptations of the classic production effect paradigm. More precisely, I will report results garnered from different types of fMRI analyses for each respective experiment: (i) conventional univariate analysis, (ii) representational similarity analysis (RSA), and (iii) pattern similarity analysis (PSA). Each type of analysis is intended to provide a different perspective on the neural processes contributing to the production effect. In brief, univariate analysis is informative as to processes shared across all stimuli in a particular condition, while RSA and PSA are both types of multivariate analysis that seek to decode (i.e., detect) known information in stimulus-specific activation patterns. I will report on results from each type of analysis in Chapters 2, 3, and 4 respectively.

The purpose of this chapter is to provide an overview of the production effect, as well as the fMRI analysis methods that will form the basis of later chapters. In Section 1.2 I will introduce the production effect and describe major theoretical issues surrounding its underlying

mechanisms. In Sections 1.3–1.7 I will outline basic principles of fMRI as well as the univariate and multivariate data analysis methods that will form the basis of this dissertation. Because Chapters 3 and 4 are concerned with task-relevant changes in multivariate decodability, Section 1.8 provides some context for how such effects can emerge in RSA and PSA. Finally, Section 1.9 lays out the specific objectives of each chapter.

1.2. The Production Effect

In the traditional production effect paradigm, participants must study two intermixed lists of words; words in one list are read aloud, words in the other are read silently. On a subsequent recognition test all of the studied words are intermixed with a number of new, unstudied words, and on each trial participants must indicate whether a presented word is "old" (i.e., they recognize it from earlier) or "new" (they do not recognize it). The production effect, coined by MacLeod et al. (2010), is the finding that participants remember a greater proportion of words read aloud compared to those read silently. This phenomenon has since been replicated in multiple studies (see Fawcett, 2013; Fawcett et al., 2023; MacLeod & Bodner, 2017 for meta-analyses and review), while variants on the original paradigm have revealed that this effect extends to types of study material other than single words, including complex sentences (Ozubko, Hourihan, et al., 2012), word-pair associations (Putnam et al., 2014), and names for presented images (e.g., Fawcett et al., 2012). The production effect also generalises to productive acts other than speaking; for example singing (Quinlan & Taylor, 2013, 2019), mouthing silently (Forrin et al., 2012; MacLeod et al., 2010; Experiment 5), and typing or writing (Forrin et al., 2012). Overall, the production effect appears to be both reliable and highly generalisable. However, its underlying cognitive mechanisms remain controversial.

1.2.1. Possible Mechanisms

Relative distinctiveness. Prevailing theoretical accounts of the production effect emphasise a role of distinctive motor and sensory experiences afforded by production. The act of vocalising a word requires the speaker to engage their jaw and tongue in an appropriate manner to produce the word's constituent phonemes in the correct order. These movements are

accompanied by sensory input—proprioceptive feedback from one's jaw and tongue, feeling one's larynx vibrate, and hearing the sound of one's own voice. These features provide a distinctive encoding experience that is specific to the word being produced; in turn, this distinctive information is thought to facilitate subsequent retrieval (Forrin et al., 2012; Jamieson et al., 2016; MacLeod et al., 2010; MacLeod & Bodner, 2017). More precisely, during encoding the sensory, motoric, and auditory features elicited by speech production are supposedly appended to the episodic representation for having read a word (the *memory trace*), making produced items more discriminable during retrieval. This is often termed *relative distinctiveness* because memory traces for aloud words are only distinctive when compared against the backdrop of non-produced words (MacLeod et al., 2010, p. 681).

Perhaps the strongest evidence for a role of relative distinctiveness is the observation that the magnitude of the production effect scales with the complexity (and therefore the distinctiveness) of the productive act. For example, a study by Quinlan and Taylor (2013) reported a gradient in recognition memory whereby singing words elicited the largest production effect, followed by reading in a loud voice, followed by reading in a normal voice (this memory advantage for singing, relative to reading aloud, was further replicated by Quinlan & Taylor, 2019). In a similar vein, less elaborate productive acts such as whispering or mouthing silently elicited a smaller production effect than reading aloud in a normal voice (Forrin et al., 2012). These findings indicate that the production effect scales with the degree of distinctiveness conferred upon study items: more elaborate productive acts entail additional distinctive elements, such as the act of modulating volume intensity (reading loudly) or modulating pitch and timbre (singing). Further supporting this conclusion is the finding that non-distinctive productive acts, such as pressing a key or speaking a non-specific word, do not elicit a production effect (MacLeod et al., 2010; Experiment 4).

Another line of evidence for relative distinctiveness (though one which has become something of a double-edged sword for its proponents) is the finding that the production effect is larger when comparing memory for aloud and silent words within subjects, as opposed to between subjects. As outlined earlier, relative distinctiveness is only relevant to memory when presented against the backdrop of non-distinct silent words (MacLeod et al., 2010, p. 681). As

such, this account predicts that the production effect should be present when memory for aloud and silent words is compared within the same subjects, but not between different subjects. While early experiments supported this prediction (e.g., MacLeod et al., 2010; Experiments 2 & 3), subsequent meta-analyses revealed a small but reliable between-subjects effect (Fawcett, 2013; Fawcett et al., 2023). Therefore, while larger within-subjects (relative to between-subjects) effects do support a role of relative distinctiveness, the reliable presence of a between-subjects effect opens the door to alternative accounts.

Strategy. A slightly different view of distinctiveness proposes that memory for aloud words may benefit from a production heuristic—that is, a metacognitive strategy that participants use to endorse or reject a presented word in a recognition task (Taikh & Bodner, 2016). This heuristic account supposes that, when presented with a previously studied word, a participant may attempt to recollect the productive experience; successful recollection allows the participant to endorse the presented word as "old": *I remember reading this word aloud, therefore I must have studied it* (MacLeod & Bodner, 2017)[1]. In a similar vein, Wakeham-Lewis et al. (2022) recently proposed a novel "sensorimotor reinstatement" account whereby at test, participants mentally simulate reading the presented word aloud. If the simulated experience matches the encoded information about the presented word, the word is endorsed as "old".

Both strategy accounts still depend on sensorimotor distinctiveness—since participants are thought to mentally replay or simulate the productive experience—however they differ from the relative distinctiveness account in two ways. First, relative distinctiveness emphasises encoding processes (the addition of sensorimotor features to the memory trace), while both strategy accounts emphasise mental processes during retrieval. Second, because strategy does not necessarily require discriminability from non-distinct items (merely successful recollection or

[1] As an aside, one may question whether this account really entails a heuristic, rather than simply the act of recollection. Here I have described a heuristic (importantly, one which is fundamentally recollective in nature), because that is how the account is generally represented in the production effect literature (e.g., Fawcett, 2013; MacLeod et al., 2010; Zhou & MacLeod, 2021).

simulation), strategy effects ought to be observable between subjects (Taikh & Bodner, 2016; Zhou & MacLeod, 2021).

While the heuristic account is often invoked in the production effect literature, and often hand-in-hand with relative distinctiveness (e.g., Forrin et al., 2012; MacLeod et al., 2010; MacLeod & Bodner, 2017; Zhou & MacLeod, 2021), to my knowledge only one study has explicitly tested it. Taikh and Bodner (2016; Experiments 3 & 4) report that a previously observed between-subjects production effect was abolished when they introduced a within-group imagery manipulation, whereby participants in both groups were instructed to visualise half the studied words' referents. These authors suggest that the introduction of the within-subjects manipulation may have caused participants to switch from using a production-based heuristic to an imagery-based one, providing tentative support for the heuristic account. These authors also found that a majority (78%) of participants subjectively report using a production heuristic in experiments where the between-subjects effect was observed (Experiment 2), however this proportion was significantly decreased (to 58%) when the effect was abolished. As for the reinstatement account; Wakeham-Lewis et al. (2022; Experiment 1) report having participants read words aloud in an unusual voice (e.g., impersonating Elvis Presley or Kermit the Frog) did not elicit a production effect for those words (while the production effect for words read in a normal voice was preserved). The authors suggest that, at test, participants usually simulate reading in their own (normal) voice; therefore the simulated experience would not match the encoded experience for unusually-voiced words. These authors further reported that allocating a larger proportion of aloud words to the unusual voice condition caused a production effect to emerge, and suggest that reading more words in unusual voices may cause participants to use those voices in their mental simulations at test (therefore endorsing unusual words) (Wakeham-Lewis et al., 2022; Experiment 2).

Attention. It has been suggested that participants allocate more attention to produced words than to nonproduced words (because the former require an overt response), and that this behaviour results in stronger encoding (Fawcett & Ozubko, 2016; Mama et al., 2018; Mama & Icht, 2019; Ozubko, Gopie, et al., 2012). This view has received some empirical support. Behavioural work has shown that disrupting attention during encoding by applying complex

auditory noise can abolish the production effect (Mama et al., 2018); while enhancing attention with a dose of methylphenidate (Ritalin) increased the magnitude of the production effect in a group of individuals with attention-deficit hyperactivity disorder (ADHD) (Mama & Icht, 2019). Taken together, these two studies indicate a relationship between participants' ability to deploy attention during encoding and the observed production effect[2]. Moreover, an electroencephalography (EEG)-based investigation of the production effect revealed that the instruction to read aloud or sing words at study elicited a P300 ERP component, relative to reading silently (Hassall et al., 2016). While the authors interpreted this as evidence of more distinctive processing (based on evidence that perceptually or semantically distinctive stimuli tend to elicit the P300, particularly in the context of encoding, e.g. Fabiani & Donchin, 1995; Karis et al., 1984) it is worth noting that the P300 has also been associated with detection of task-relevant stimuli (e.g., Dinteren et al., 2014 for review). As such, the results of Hassal et al (2016) are arguably consistent with an attentional account whereby aloud words are attended to because they are task-relevant.

Semantics. Recognition memory for aloud words may also benefit from elaborate semantic processing during encoding. This idea was proposed by Fawcett et al. (2022), who reported that the production effect was diminished in a two-alternative forced choice task when non-target words (lures) were semantically related to (previously studied) target words. These authors suggested that, in contexts where semantic lures are absent, participants may rely on encoded semantic information for recognition of aloud words. More precisely, reading words aloud might facilitate access to semantic features of the presented words; these features may be appended to the episodic representation for that word. In turn, the appended semantic features may to some extent guide recognition decisions at test—this would explain why semantic lures disrupted recognition performance for aloud words, because the (ordinarily useful) semantic features would not be diagnostic when choosing between semantically related alternatives. The notion that semantic processing might underlie the production effect is a relatively new idea; to

[2]It is worth noting that both of these studies examined the production effect using a free recall procedure. It is not clear if these results are applicable to the production effect in recognition memory.

my knowledge, no other studies have investigated this possibility. Nevertheless, it presents an interesting avenue for further research.

Summary. Multiple mechanisms—relative distinctiveness, strategy, attention, and semantics—have been proposed to explain the production effect. Although distinctiveness appears to dominate the literature (indeed, it seems impossible to discuss the production effect without reference to distinctiveness), the reliable presence of a between-subjects effect necessitates different explanations. Attentional, semantic, and strategy-based accounts have arisen to meet this demand, and there appears to be at least some support for each. Importantly, the mechanisms discussed in this section are not necessarily mutually exclusive, and it is entirely possible that each contributes somewhat to the production effect (an idea expressed by multiple authors, e.g., Fawcett, 2013; Fawcett et al., 2022; Ozubko, Gopie, et al., 2012). Contributions from different processes/mechanisms likely depend on distinct neural correlates. This possibility may be explored through univariate and multivariate analysis of fMRI data, discussed below.

1.3. Functional Magnetic Resonance Imaging (fMRI)

fMRI is a neuroimaging technique that measures neuronal activity indirectly by detecting changes in Blood Oxygen Level Dependent (BOLD) contrast, which in turn is determined by the relative concentration of oxygen-bound haemoglobin (oxyhaemoglobin) to oxygen-unbound haemoglobin (deoxyhaemoglobin) in the brain's blood vessels. In brief, these two forms of haemoglobin have different magnetic properties, such that the recorded MRI signal is relatively stronger in areas where there is a higher concentration of oxyhaemoglobin (Ogawa et al., 1990). This property of the BOLD signal is relevant to neuronal activity because neurons require oxygen to perform action and synaptic potentials (the basic physiological processes underlying brain activity). Neural activation—that is, a transient and synchronous increase in activity of many adjacent neurons in response to a stimulus—is therefore characterised by a large local influx of oxyhaemoglobin to the activated area (Malonek & Grinvald, 1996). As such, relative changes in the ratio of oxy- to deoxy-haemoglobin within a patch of cortex provide an indirect measure of neuronal activity in that area. The BOLD signal is recorded as a continuous time series at each of

a number of voxels (volumetric pixels), which in turn make up a three-dimensional image of the brain. The spatial resolution of fMRI is therefore limited by the size of these voxels; typically, a functional image will have a resolution of a few millimetres, meaning that a single voxel will encompass tissue containing thousands of neurons. Having said this, the spatial resolution of fMRI is relatively high compared to other imaging techniques, therefore it currently represents the gold standard for non-invasive spatial localisation of brain function. By contrast, the temporal resolution of fMRI is very poor; the hemodynamic response function (HRF) - which describes the canonical time course of the BOLD response - takes approximately 6-10 seconds to peak following stimulus presentation, and only returns to baseline after approximately 15 seconds (e.g., Arichi et al., 2012).

The BOLD signal at any given voxel is ultimately arbitrary; making meaningful inferences about this signal depends on identifying systematic variance associated with experimental manipulations. In brief, the continuous BOLD time course may be compared to a model time course in which temporally-defined events (e.g., periods in which stimuli are presented) are convolved with HRFs. Such a model time course represents a sensible prediction about the time course of the BOLD response in a given voxel, assuming that the voxel in question is sensitive to (i.e., activated by) the presented stimuli. Therefore, by comparing the observed data to the model time course at every voxel, one may estimate each voxel's sensitivity to stimuli in a given experimental condition.

1.4. Univariate Analysis of fMRI Data

To date, most fMRI research has depended on univariate analysis methods. Fundamentally, these methods consider each voxel as an independent unit of measurement and are contrastive in nature. In a typical univariate design, at each voxel, the activation elicited by a given experimental condition (that is, the aggregate response to all stimuli within that condition) may be contrasted either with baseline periods or with activation elicited by other experimental conditions. While these contrasts are, strictly speaking, performed at each individual voxel, in practice inferences are made at the level of clusters. Clusters, in turn, are defined as groups of contiguous, statistically significant voxels (that is, whose activation is significantly different from

baseline and/or between conditions) which meet some other a-priori criteria such as spatial or statistical extent, or a combination of the two (S. M. Smith & Nichols, 2009).

Most univariate-based research to date has been concerned with neural processes at the level of broad experimental conditions. Here, researchers assume that neural responses that are consistently elicited by different exemplars of a given condition—which might be a task or stimulus category—are representative of that condition broadly. Any variance between stimuli are therefore considered unimportant and disregarded as noise. For example, one might present participants with images of different faces (faces condition) and images of different household objects (objects condition). By contrasting activation between these two conditions, one may identify areas that are relatively more sensitive to faces than to objects. One might even determine that a particular area is specialised for a specific domain, such as face processing, by demonstrating that the area in question responds preferentially to images of faces compared with multiple other image classes (such as houses, tools, and other body parts, e.g., Kanwisher et al., 1997). Critically, information revealed by such contrasts reflect processes that are common to all stimuli in the condition(s) being examined.

The contrastive approach described above has formed the backbone of much of cognitive neuroscience over the past two decades. However, this approach is limited in two major ways. First, univariate analyses only reveal focal clusters of voxels that are maximally responsive to a particular condition; in reality, stimulus- and task-relevant processes are widely distributed amongst less-responsive voxels as well (Haxby et al., 2001). Second, the inherent assumption that voxels represent independent measurement units means that univariate analyses are relatively insensitive to fine-grained differences in neural responses between unique stimuli. This is because the activation level of a single voxel (or, indeed, the isolated activation levels in many contiguous voxels) is minimally informative as to a participant's perceptual or cognitive state, considering that most stimuli elicit complex responses spanning hundreds or thousands of voxels. These issues fundamentally limit our ability to characterise neural responses to unique stimuli using univariate analyses.

1.5. Multi-Voxel Pattern Analysis (MVPA)

Multivariate approaches to analysing fMRI data—collectively referred to as multi-voxel pattern analyses (MVPA)—can to some extent overcome the limitations of univariate analyses outlined above. Broadly speaking, MVPA approaches characterise the neural response to a stimulus as a pattern of activation across *N* contiguous voxels. Many applications of MVPA are decoding-based; that is, they aim to predict stimulus- or experimentally-defined information from observed activation patterns. Decoding may be conceptualised broadly as a means of describing statistical dependency between observed activation patterns and known information (Kriegeskorte & Kievit, 2013). Known information might refer to broad stimulus categories (e.g., in the case of a binary classifier), variance along a particular stimulus dimension (as in RSA), or stimulus identity (as in PSA). I will elaborate on the latter two cases in Sections 1.6 and 1.7 respectively.

From a decoding perspective, the benefits of examining activation patterns (as opposed to activation levels in individual voxels) are twofold. First, the distributed pattern of activation elicited by a particular condition or stimulus often contains information that is unique to that condition and, in turn, discriminable from activity elicited by other conditions (M. Peelen & Downing, 2022). This has led to the finding that neural responses are distributed across relatively large, often overlapping areas of cortex (Haxby et al., 2001)—contrary to the modular view provided by univariate analyses that specific patches of cortex are dedicated to processing specific types of stimuli (e.g., Kanwisher et al., 1997). A second benefit is that individual stimuli can often be differentiated from one another based on the unique patterns of activation that they elicit, regardless of the stimulus class or experimental condition to which they belong. For example, a number of studies have shown that images of unique objects elicit discriminable patterns of activation in ventral temporal cortex (e.g., see Grill-Spector & Weiner, 2014 for review), indicating that this area processes visual information in an object-specific manner.

An important concept for any application of MVPA is that pairs of activation patterns can be quantitatively compared. Activation patterns may be represented mathematically as *pattern vectors*—that is, numerical sequences of length *N* where *N* is the number of voxels from which the pattern was extracted, and each value in the sequence describes the activation level in the

corresponding voxel. This framework makes pairwise comparisons of activation patterns mathematically trivial, provided that both patterns are derived from an equal number of voxels. Different approaches to comparing patterns may use measures of similarity or dissimilarity. Measures of pattern *similarity* quantify the degree to which pairs of patterns are alike; for example, based on their Pearson correlation. By contrast, measures of *dissimilarity* quantify the degree to which patterns are different (e.g., correlation distance; 1 – Pearson correlation). Measures of similarity and dissimilarity form the basis of many decoding-based applications of MVPA. Two such applications are RSA and PSA, discussed below.

1.6. Representational Similarity Analysis (RSA)

RSA is a method which enables quantitative comparison of stimulus-specific information garnered from different types of measurement, including measures of neural activity patterns (e.g., from fMRI), measures along a particular stimulus dimension, or formal computational models (Kriegeskorte et al., 2008). Central to RSA is the representational dissimilarity matrix (RDM), which is a matrix comprising pairwise dissimilarity values for a given set of stimuli. An RDM may be derived from any of the types of measurement listed above. Here I will focus on measures of stimulus properties as these are relevant to the analyses undertaken in Chapter 3.

Any stimulus can be represented in multiple dimensions. For example, a printed word may be described in terms of its visual properties, its meaning (semantics), or the sound of it being spoken (phonology). Assuming that one has a meaningful way of quantifying the (dis)similarity between items with respect to these properties, one can build a corresponding RDM based on dissimilarity values between each pair of stimuli; one can therefore construct separate RDMs to represent visual properties, semantic properties, and so on. RDMs are said to characterise the *representational geometry* of a stimulus set—that is, a model of the information content that is based on relationships between stimuli, according to the measure being examined (Kriegeskorte & Kievit, 2013).

The value of representing information content in this way is that it provides a valid means of assessing correspondence (i.e., representational similarity) between different sources

(Kriegeskorte et al., 2008). For example, one might assess how well neural activity patterns in a specific path of cortex reflect visual information by simply correlating a neural RDM (derived from fMRI data) with a visual RDM (derived from pixel-wise RGB values) (Kriegeskorte & Kievit, 2013) or by implementing a general linear model that includes a neural RDM as the dependent variable and a visual RDM as a regressor (Oosterhof et al., 2016). RSA therefore allows us to decode different kinds of information—such as information about specific stimulus properties—in neural activation patterns. Within this framework, it is useful to describe RDMs that reflect specific stimulus properties as *hypothesis models*. One might hypothesise that visual information will be decodable in a particular brain area (e.g., primary visual cortex); a visual RDM would therefore represent a formal—and testable—model of one's prediction about the information content of neural activity patterns in the area in question.

Much prior work has used RSA to explore the decodability of various types of information in neural data. For example, low-level visual information has been decoded from activity elicited in early visual cortex when viewing complex pictures or words (e.g., when visual information is defined by binary silhouette models; Devereux et al., 2013; Kriegeskorte et al., 2008). Semantic information, such as animal or object category distinctions, has been decoded in ventral temporal cortices (Borghesani et al., 2016; Devereux et al., 2013; Nastase et al., 2017; Tong et al., 2022; Wang et al., 2018); while phonological and orthographic information have both been decoded in the left fusiform gyrus (Fischer-Baum et al., 2018; Li et al., 2022; Qu et al., 2022; Zhao et al., 2017), which notably houses the visual word-form area (VWFA), thought to play a critical role in reading (see Dehaene & Cohen, 2011 for review).

Overall, RSA is a powerful tool that allows us to decode different kinds of information from neural activity patterns. The value of RSA in the context of this dissertation is that it can be used to compare the decodability of specific types of information that are relevant to word reading (e.g., phonology, semantics, etc.) between different tasks, such as reading aloud and reading silently. Such task-related differences in decodability will form the basis of Chapter 3.

1.7. Pattern Similarity Analysis (PSA)

In contrast to RSA, which is concerned with mutual information between different sources (e.g., stimulus properties and fMRI response patterns), PSA is directly concerned with relationships between pairs of neural activation patterns. As an aside, the distinction between these two terms is not always clear in MVPA literature—for example, applications of PSA (as conceptualised here) are sometimes labelled as RSA. For convenience I use RSA to describe the process of comparing neural data with a given hypothesis model(s), as described in the previous section. By contrast, I consider PSA to encompass simple comparisons between pairs of activation patterns, as described below. Many studies employing PSA have focussed on episodic memory; as such, the following description of PSA is framed in terms of this research area.

As outlined in Section 1.5, the similarity of two neural pattern vectors (i.e., neural responses from a set of voxels) may be quantified by their Pearson correlation. This simple principle has been leveraged to great effect in research examining neural correlates of episodic memory. Such research has identified two distinct processes that are relevant to episodic encoding and retrieval: *neural pattern reactivation* and *neural pattern transformation*.

1.7.1. Neural Pattern Reactivation

Neural pattern reactivation refers to recapitulation of cortical states across multiple presentations of the same item. I use the term *item* here instead of *stimulus*, because the same item may be represented in different stimulus formats (a printed word, a picture, etc.) across repetitions. A robust finding in MVPA studies is that pairs of responses elicited by the same item are more similar than pairs elicited by different items (Favila et al., 2018; Hasinski & Sederberg, 2016; Kuhl & Chun, 2014; Long & Kuhl, 2021; Ritchey et al., 2013; Xiao et al., 2017; Xue et al., 2010, 2013; Zeithamova et al., 2017). In such cases, we can say that item identity is decodable in the observed activation patterns, because the relationships between different patterns can be differentiated by (i.e., are statistically dependent upon) known information—that is, the identity of the eliciting stimulus. The decodability of item identity is thought to be relevant to encoding and retrieval because it often tracks with subsequent memory success. That is, the fidelity of

responses to a unique study item, either across encoding episodes (Hasinski & Sederberg, 2016; Xue et al., 2010, 2013; Zeithamova et al., 2017), or between encoding and retrieval (Davis et al., 2014; Ritchey et al., 2013) is often higher for subsequently remembered items compared to forgotten items. Similar memory-dependent reactivation effects have been reported from experiments where participants study stimulus pairs (picture-picture or picture-word pairings) during encoding, and later are presented with a single item and must retrieve its paired associate (Favila et al., 2018; Jonker et al., 2018; Kuhl et al., 2013; Kuhl & Chun, 2014; Staresina et al., 2012; Tompary et al., 2016; Trelle et al., 2019; Wing et al., 2015).

We can, to some extent, make inferences about the information content of these reactivated patterns based on known functional properties of cortical areas in which reactivation is observed. Reactivation in ventral and lateral temperooccipital cortices (Danker et al., 2017; Hasinski & Sederberg, 2016; Long & Kuhl, 2021; Ritchey et al., 2013; Wing et al., 2015; Xue et al., 2010; Zeithamova et al., 2017) may reflect reinstatement of visual perceptual processes, as both of these areas have long been associated with visual object recognition (e.g., Goodale & Milner, 1992; Grill-Spector et al., 2001; M. V. Peelen & Downing, 2017). By contrast, reactivation in frontoparietal (Kuhl & Chun, 2014; Xiao et al., 2017; Xue et al., 2010, 2013; Zeithamova et al., 2017) and medial temporal (Davis et al., 2014; Staresina et al., 2012; Tompary et al., 2016) areas may reflect amodal (i.e., perceptually invariant) processing demands that were present during encoding. Medial temporal lobe structures (particularly the hippocampus) have long been understood as essential for successful episodic encoding (e.g., Squire, 2009), while frontoparietal cortices have been implicated in higher-level executive processes relevant to encoding and retrieval (Ciaramelli et al., 2008). That being said, some studies have revealed that visual category information can be decoded from activity patterns in frontoparietal regions (e.g., Kuhl et al., 2013; Long & Kuhl, 2021; Zeithamova et al., 2017), indicating that these areas may represent specific mnemonic content.

Overall, reactivation has been revealed as a robust neural substrate of successful remembering. This observation is consistent with the view that successful remembering should, at least to some extent, entail reinstatement of cognitive operations (and their underlying neural states) that were elicited during encoding.

1.7.2. Neural Pattern Transformation

In contrast to reactivation, neural pattern transformation refers to systematic alteration of activation patterns (Favila et al., 2020; Xue, 2022). Transformation has been operationalised in a few different ways, but all broadly describe changes in decodability (of both item identity and stimulus properties) between encoding and retrieval. On the one hand, transformation may be captured as a shift in the spatial position of encoded information. Xiao et al. (2017) and Favila et al. (2018) both report that during encoding, item identity was decodable in ventral visual areas but not (or was markedly less so) in frontoparietal areas; during retrieval, however, these effects was reversed such that item identity was more decodable in frontoparietal areas. Favila et al. (2018) additionally reported that the decodability of stimulus colour and taxonomic class (properties that were both dissociable from item identity) was similar in ventral visual areas across encoding and retrieval, but was significantly higher during retrieval (compared to encoding) in lateral parietal cortices. These findings suggest that stimulus information, initially encoded in ventral visual areas, may be reinstated in frontoparietal areas during retrieval. Moreover, Xiao et al. (2017) reported that the representational geometry elicited in ventral visual cortex during encoding was significantly correlated with that of frontoparietal areas during retrieval, thus providing an explicit link between encoded information in one area with retrieved information in another.

Transformation may also be captured as a systematic change in decodability within one or more discrete region(s), without necessarily invoking corresponding changes in other regions. Importantly, in order to be considered systematic, such within-region transformation should not reflect mere a weakening of response patterns or an increase in noise between encoding and retrieval (either of which would result in decreased decodability). The findings of Xiao et al. (2017) and Favila et al. (2018) arguably provide examples of such within-region transformation: in both cases, stimulus information was more decodable in frontoparietal cortices during retrieval compared to encoding, suggesting a systematic change in how stimuli were represented in those areas. Furthermore, within-region transformation was explicitly demonstrated by Chen et al. (2017) using a novel between-subjects analysis. Chen et al. (2017) had participants watch

(encode) and later recall information about a movie, and examined activation patterns elicited by discrete scenes during both tasks. These authors correlated activation patterns for the same scenes across participants: correlations were performed either between encoding and recall patterns (encoding-recall correlations) or between recall patterns (recall-recall correlations). In a number of areas, notably including frontal and parietal cortices, recall-recall correlations were significantly higher than encoding-recall correlations, suggesting that neural responses to scenes changed systematically between encoding and retrieval. As before, these changes could not reflect mere weakening of encoded patterns, because item-specific information was robustly decodable during recall. This within-region approach to characterising transformation is particularly relevant to my dissertation because it is the approach that I will adopt in Chapter 4.

It has been suggested that neural pattern transformation (either within or across regions) reflects systematic reorganisation of encoded information, perhaps as part of the encoding process and/or to meet specific cognitive demands involved in retrieval (Favila et al., 2020; Xue, 2022; see Section 1.8 for a discussion of task-dependent transformation). Such reorganisation might entail semanticization of encoded perceptual information - that is, an emphasis on semantic features of the encoded event, as opposed to low-level sensory properties. In support of this, some electrophysiology (i/EEG)-based work has linked neural transformation to a shift from perceptual to semantic decodability between encoding and retrieval. Linde-Domingo et al. (2019) report that neural responses to presented images of objects could be classified based on perceptual stimulus distinctions (grayscale vs colour) earlier in time after stimulus presentation compared to semantic distinctions (animate vs inanimate) during encoding; this effect was reversed during retrieval. Moreover, an RSA by Liu et al. (2021) revealed that successful retrieval of word-picture pairs was associated with greater semantic, relative to perceptual, decodability during encoding. Both of these studies provide evidence that encoding and retrieval are associated with differential decodability of perceptual versus semantic information.

Unlike research surrounding reactivation, few studies (to my knowledge) have explicitly linked transformation effects to behavioural retrieval success. However, two data points stand out. First, there is Liu and colleagues' (2021) observation that subsequently remembered items were associated with greater decodability of semantic information. Moreover, Chen et al. (2017)

reported that the magnitude of event-specific transformation in medial parietal cortex was significantly correlated with the number of participants who successfully remembered that event (see J. Chen et al., 2017, Supplementary Figure 9). These findings, together with the general observation that transformation effects are often identified between encoding and retrieval, suggest that neural pattern transformation represents an important component of successful remembering. While still a relatively new area of research, transformation as a component of retrieval is consistent with the traditional view of remembering as a reconstructive (rather than reproductive) act (Schacter et al., 1998).

1.8. Task Dependent Decodability

An important finding from both RSA and PSA studies is that the decodability of information from activation patterns is often dependent on participants' internal cognitive state, which often varies systematically according to task-related cognition or response goals. This concept is relevant to my dissertation, which aims to investigate differences in decodability between aloud and silent reading. A convenient framework for understanding such task-based effects is that of embodied and grounded cognition. Proponents of this framework argue that our internal cognitive states, and in turn the neural substrates of those states, are influenced (if not determined) by the manner in which we receive input from and interact with our environment (Matheson & Barsalou, 2018). A natural prediction of this perspective is that the decodability of information in neural responses, elicited by a given task, will be driven in large part by an individual's experience of that task—both in terms of explicit sensory input and top-down, goal-directed cognition. Accordingly, any hypotheses that we make about neural states (e.g., during aloud versus silent reading) should take such experiential processing into account.

An excellent example of how task-related cognition can affect decodability in RSA comes from Nastase et al. (2017). These authors presented participants with video clips of animals behaving in their natural environment, and participants were instructed to make decisions concerning either the animals' taxonomy or their behaviour. When attending to taxonomy, participants' responses in ventral temporal cortex showed a higher correlation with a

taxonomically-defined hypothesis model than with a behavioural model; when attending to behaviour, however, this effect was reversed within the same brain region. Similar effects have been reported in RSA studies of single word reading, which have revealed task-dependent variability in taxonomic versus contextual decodability in VWFA (Wang et al., 2018), as well variable phonological and orthographic decodability when reading aloud versus a perceptual judgement task (Qu et al., 2022). These findings indicate that goal-directed cognitive states may modulate the decodability of particular stimulus features.

Neural reactivation is also task-dependent. Of particular interest here are studies in which participants study arbitrary word-picture pairs and are then instructed to recall the picture when cued with its paired word. Such studies often report decodability of (picture-specific) object identity information in lateral/ventral visual and lateral frontoparietal areas—which are notably associated with object-specific visual processing (see end of Section 1.7.1)—during retrieval (Favila et al., 2018; Jonker et al., 2018; Kuhl et al., 2013; Kuhl & Chun, 2014; Staresina et al., 2012; Tompary et al., 2016; Trelle et al., 2019; Wing et al., 2015). The fact that object identity information is still detectable in these areas during retrieval, despite the pictures depicting those objects being perceptually absent, indicates that successful retrieval entails reinstatement of processes that were present during encoding. Stated differently, one's memory for presented objects appears to be grounded in the perceptual experiences through which those objects were encoded.

With respect to neural transformation, the demands imposed by a particular retrieval task may necessitate reorganisation of the encoded information in a manner that best suits that task. This principle was demonstrated by Favila et al. (2018; Experiment 2), who reported variable decodability of stimulus colour or taxonomic class in lateral parietal cortex when participants were instructed to emphasise each of these respective features during retrieval. This finding demonstrates a direct relationship between top-down retrieval goals and the decodability of perceptual information. Moreover, one might consider encoding and retrieval to involve different goals generally (perception versus mental reconstruction), which may account for the emergence of object identity or perceptual information in different cortical areas during encoding versus retrieval (J. Chen et al., 2017; Favila et al., 2018; Xiao et al., 2017).

In summary, the decodability of information as revealed by RSA and PSA is dependent on both perceptual experiences and goal-directed cognition. This is relevant to my dissertation because, as I will argue throughout, aloud and silent reading represent fundamentally different perceptual and cognitive states. Therefore, it is entirely possible that the production effect—defined as a contrast between these tasks—may entail differential decodability of perceptual and/or cognitively relevant information.

1.9. Goals Of The Current Work

As stated earlier, the overall goal of this dissertation is to examine neural correlates of the production effect. Below I outline specific research questions that will be empirically tested in Chapters 2, 3, and 4. These research questions are naturally driven by theoretical accounts presented in Section 1.2.1. To be clear, it is not my intention to explicitly test or evaluate these accounts; rather, I consider them useful for making predictions and inferences about observed results.

In Chapter 2 I aim to identify brain areas that broadly contribute to the production effect—that is, areas which show effects common to all stimuli in the aloud reading condition versus silent reading. fMRI analyses will entail univariate contrasts between words that were read aloud versus silently, during a study phase (encoding) and test phase (retrieval). Based on the view that reading aloud entails rich sensorimotor and auditory experiences, and that these experiences facilitate later retrieval, reading aloud should elicit more activation (relative to silent reading) in areas governing those processes during study and test. Alternatively, aloud reading may be associated with greater activation in hubs for semantic (e.g., ventral temporal) or attentional (e.g., frontoparietal) processes.

In Chapter 3 I will use RSA to examine the decodability of different kinds of information (that are relevant to single word reading) during aloud versus silent reading. Aloud and silent reading entail different response goals, sensory experiences, and (perhaps) semantic and attentional processes. Therefore, these two tasks arguably reflect fundamentally different perceptual and cognitive states. Based on this reasoning, they may entail differential decodability

of different kinds of information (visual, orthographic, phonological, etc.) during encoding. I will apply RSA to fMRI data acquired while participants read single words either aloud or silently, with a view to ascertain which (if any) types of information are more decodable during aloud versus silent reading.

In Chapter 4 I will use PSA to examine relative contributions of neural reactivation and transformation to successful recognition of words that were previously read aloud or silently. I will compute behaviourally-relevant (that is, contrasts of remembered and forgotten items) measures of reactivation and transformation, based on the similarity or divergence of activation patterns between encoding and retrieval. Both processes are arguably compatible with theoretical accounts of the production effect. For example, a role distinctiveness during retrieval might manifest as reactivation in sensorimotor cortices; alternatively, evaluative recognition judgements (e.g., "do I remember reading this word aloud?") may entail transformation of encoded perceptual information. Regardless, differential contributions of either process for aloud versus silent words would signal a meaningful neural substrate of the production effect.

CHAPTER 2. NEURAL CORRELATES OF THE PRODUCTION EFFECT: AN FMRI STUDY

2.1. Publication Information

Bailey, L. M., Bodner, G. E., Matheson, H. E., Stewart, B. M., Roddick, K., O'Neil, K., Simmons, M., Lambert, A. M., Krigolson, O. E., Newman, A. J., & Fawcett, J. M. (2021). Neural correlates of the production effect: An fMRI study. Brain and Cognition, 152, 105757. https://doi.org/10.1016/j.bandc.2021.105757[3]

2.2. Abstract

Recognition memory is improved for items produced at study (e.g., by reading them aloud) relative to a non-produced control condition (e.g., silent reading). This production effect is typically attributed to the extra elements in the production task (e.g., motor activation, auditory perception) enhancing item distinctiveness. To evaluate this claim, the present study examined the neural mechanisms underlying the production effect. Prior to a recognition memory test, different words within a study list were either read aloud, silently, or while saying "check" (as a sensorimotor control condition). Production improved recognition, and aloud words yielded higher rates of both recollection and familiarity judgments than either silent or control words. During encoding, fMRI revealed stronger activation in regions associated with motor, somatosensory and auditory processing for aloud items than for either silent or control items. These activations were predictive of recollective success for aloud items at test. Together, our findings are compatible with a distinctiveness-based account of the production effect, while also pointing to the possible role of other processing differences during the aloud trials.

2.3. Statement of Student Contributions To Manuscript

My contributions to this chapter included writing the published manuscript (I fully revised an earlier unpublished draft), data curation, fMRI data analysis and interpretation, and creation of tables and figures. The only major element of this manuscript that I did not contribute was behavioural data analysis (performed by one of my co-authors, Jonathan Fawcett).

[3] This manuscript has been reformatted for integration into this thesis: page and section numbers are not as they appear in the published manuscript. An additional section ("Additional contrasts – results & discussion") has been added to the supplementary materials that were originally published alongside the article.

Brain and Cognition 152 (2021) 105757

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Neural correlates of the production effect: An fMRI study

Lyam M. Bailey [a], Glen E. Bodner [b], Heath E. Matheson [c], Brandie M. Stewart [a], Kyle Roddick [a], Kiera O'Neil [a], Maria Simmons [a], Angela M. Lambert [d], Olave E. Krigolson [e], Aaron J. Newman [a], Jonathan M. Fawcett [f,*]

[a] Dalhousie University, Department of Psychology and Neuroscience, Halifax, NS B3H 4R2, Canada
[b] Flinders University, College of Education, Psychology and Social Work, Adelaide, SA 5001, Australia
[c] University of Northern British Columbia, Psychology Department Prince George, BC V2N 4Z9, Canada
[d] University of Calgary, Department of Psychology, Calgary, AB T2N 1N4, Canada
[e] University of Victoria, School of Exercise Science, Victoria, BC V8W 2Y2, Canada
[f] Memorial University of Newfoundland, Department of Psychology, St. John's, NL A1B 3X9, Canada

ARTICLE INFO

Keywords:
Language
Speech
Memory
Brain
Human
Neuroimaging
Encoding
Recognition
Multivoxel pattern analysis
Representational similarity analysis

ABSTRACT

Recognition memory is improved for items produced at study (e.g., by reading them aloud) relative to a non-produced control condition (e.g., silent reading). This *production effect* is typically attributed to the extra elements in the production task (e.g., motor activation, auditory perception) enhancing item distinctiveness. To evaluate this claim, the present study examined the neural mechanisms underlying the production effect. Prior to a recognition memory test, different words within a study list were read either aloud, silently, or while saying "check" (as a sensorimotor control condition). Production improved recognition, and aloud words yielded higher rates of both recollection and familiarity judgments than either silent or control words. During encoding, fMRI revealed stronger activation in regions associated with motor, somatosensory, and auditory processing for aloud items than for either silent or control items. These activations were predictive of recollective success for aloud items at test. Together, our findings are compatible with a distinctiveness-based account of the production effect, while also pointing to the possible role of other processing differences during the aloud trials as compared to silent and control.

Introduction

A central issue in memory research is understanding how encoding strategies influence subsequent retention. An encoding strategy that has shown great promise for improving memory is the simple act of reading items aloud rather than silently. The memory advantage for reading aloud has recently been termed the "production effect" (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; see also MacLeod & Bodner, 2017), and this advantage has been found for variants of production including mouthing, typing, writing, and spelling (e.g., Fawcett, Quinlan, & Taylor, 2012; Forrin, MacLeod, & Ozubko, 2012; MacLeod et al., 2010), singing (Quinlan & Taylor, 2013; Hassall, Quinlan, Turk, Taylor, & Krigolson, 2016), and even drawing (Wammes, Meade, & Fernandes, 2016). The production effect often scales up with the complexity of the productive act (e.g., Fawcett et al., 2012; Forrin et al., 2012; Quinlan & Taylor, 2013). Further, an influence of production is seen on various tests of explicit long-term memory including recognition and recall (e.g., Conway & Gathercole, 1987; Fawcett et al., 2012; Lin & MacLeod, 2012), but not on tests of implicit memory (MacLeod et al., 2010). This effect is typically larger when manipulated within-subjects as opposed to between-subjects (Bodner, Taikh, & Fawcett, 2014; Fawcett, 2013; Fawcett & Ozubko, 2016), suggesting that context modulates its influence (MacLeod et al., 2010; Ozubko, Gopie, & MacLeod, 2012).

The production effect has most often been attributed to production enhancing the distinctiveness of items in memory (e.g., Conway & Gathercole, 1987; Dodson & Schacter, 2001; MacLeod et al., 2010). Silent reading invokes both orthographic (visual) and semantic (meaning) processing but reading aloud necessitates additional productive elements — including engagement of the articulatory-motor system followed by auditory perception of the spoken item. The *production record* laid down at encoding thus includes these additional elements (Fawcett, 2013), which can later serve to help retrieve items from memory.

* Corresponding author.
E-mail address: jfawcett@mun.ca (J.M. Fawcett).

https://doi.org/10.1016/j.bandc.2021.105757
Received 11 May 2020; Received in revised form 16 May 2021; Accepted 18 May 2021
Available online 12 June 2021

By one version of a distinctiveness account, participants employ a distinctiveness heuristic at test (Dodson & Schacter, 2001; MacLeod et al., 2010) whereby access to the production record is used to discriminate studied from non-studied items ("If I can recollect saying it aloud at study, it was studied"). Although participants often report using this strategy (Fawcett & Ozubko, 2016), recent computational modeling of the production effect suggests that use of the production record need not be intentional/conscious (Jamieson, Mewhort, & Hockley, 2016). Instead, Jamieson et al. suggest that a distinctiveness account of the production effect may reflect intrinsic retrieval dynamics favouring recovery of items containing discriminative features.

Initial failures to observe a between-subjects production effect were taken as important evidence in favour of a distinctiveness account (e.g., MacLeod et al., 2010; Ozubko et al., 2012). To the extent that items are thought to be "distinctive" only in relation to other "non-distinctive" items from the same list (Hunt, 2006), a relative distinctiveness account predicts a production effect in within-subject designs but not in between-subject designs where there is no "backdrop" of non-produced items against which produced items stand out. However, meta-analyses and subsequent experiments have revealed a between-subject production effect in recognition memory (e.g., Bodner et al., 2014; Fawcett, 2013; Fawcett, Baldwin, Drakes, & Willoughby, submitted for publication; Forrin & MacLeod, 2016). Furthermore, Fawcett and Ozubko (2016) showed that the within-subject production effect reflects an increase in both familiarity (i.e., a sense of "knowing" that the item had been studied) and recollection (i.e., the ability to re-experience the episode in which the item had been studied) for produced items. In contrast, the between-subject production effect reflects only an increase in familiarity. These findings are difficult to explain with reference to distinctiveness alone and suggest that additional factors may contribute to the within-subject effect (Fawcett & Ozubko, 2016). For example, participants self-report paying more attention to the aloud items in post-experimental questionnaires (Fawcett & Ozubko, 2016) and are less likely to mind-wander when reading aloud than reading silently (Varao Sousa, Carriere, & Smilek, 2013).

The neural basis of speech production and human memory

The present study is the first to use functional magnetic resonance imaging (fMRI) to isolate the brain regions and processes contributing to the read-aloud version of the production effect. To inform this work, we first summarize research characterizing the neural networks involved in the processes invoked during a production task, including single-word speech production and memory retention.

Neuroimaging work in the area of speech production has implicated a left-lateralized network including the inferior frontal gyrus (IFG), middle frontal gyrus (MFG), superior frontal gyrus (SFG)—especially in the supplementary motor area (SMA), anterior cingulate cortex (ACC), insula, intra-parietal sulcus (IPS) and adjacent superior parietal lobule (SPL), angular gyrus (AG), and occipito-temporal cortex (including the fusiform, inferior occipital, middle occipital, and inferior temporal gyri; for meta-analyses, see Martin, Schurz, Kronbichler, & Richlan, 2015; Taylor, Rastle, & Davis, 2013; Vigneau et al., 2006; Vigneau et al., 2011; Wagner, Sebastian, Lieb, Tuscher, & Tadic, 2014). Subcortical regions have been similarly implicated, including the caudate nucleus, putamen, and thalamus. Right hemisphere brain regions are also consistently activated – albeit on a more restricted basis – including the IFG, pre-central gyrus (premotor cortex), middle temporal gyrus, and inferior parietal lobe.

Beyond production itself, participants also maintain – and further process – studied items in working memory (WM). Meta-analyses of studies involving such processes highlight a network of regions primarily in frontal and parietal cortices (Nee et al., 2012; Owen, McMillan, Laird, & Bullmore, 2005; Rottschy et al., 2012). Greatest convergence is observed in the superior frontal sulcus (separating the SFG and MFG) and superior parietal lobule, regions that relate most strongly to executive function. Additional regions implicated include the IFG and MFG (associated with maintenance of verbal information, and selection of information), ACC (e.g., task switching), and inferior parietal lobe (directing attention between items; Nee et al., 2012; Owen et al., 2005; Rottschy et al., 2012).

Work on longer-term memory representations has often focused on the medial temporal lobe (MTL), encompassing the perirhinal cortex, parahippocampal cortex, entorhinal cortex, and hippocampus. Activation levels in the hippocampus, left IFG/MFG, bilateral regions of premotor cortex, IPS/SPL, fusiform cortex, and hippocampus have been identified as predictive of later memory performance (Kim, 2011). Tasks involving item-specific encoding were associated with stronger activation of posterior IFG/MFG and premotor cortex, and the IPS/SPL. In contrast, associative encoding (i.e., memory for items in relation to one another) was more strongly related to activity in anterior IFG/frontal pole, insula, and hippocampal regions. Premotor and posterior parietal activation may relate to increased attention to individual items during such tasks (Kim, 2011), in line with the suggestion that parietal regions act as an "attentional circuit-breaker" that re-orients attention to relevant stimuli (Astafiev, Shulman, & Corbetta, 2006; Corbetta & Shulman, 2002; Vilberg & Rugg, 2008).

Recent work has highlighted the contrast between inferior and superior parietal regions in memory retrieval. For example, the Attention to Memory (AtoM) model (Ciaramelli, Grady, & Moscovitch, 2008) draws analogies between (1) bottom-up, alerting processes in attention and attentional capture by retrieved memory items, supported by the inferior parietal lobe (IPL); and (2) the SPL's involvement in top-down strategic orienting of attention and more effortful memory retrieval (e.g., of low-confidence items). A more recent model (Sestieri, Shulman, & Corbetta, 2017) takes a slightly different view, associating the SPL/IPS region with maintenance and attentional selection of task-relevant information retrieved from memory, and the IPL (primarily AG) region with recollecting specific details of an event or other retrieved information.

In addition to areas involved in encoding and retrieval generally, a popular view holds that successful retrieval should involve at least partial reinstatement (or recapitulation) of neural states that were present at encoding. Indeed there is evidence that activation during retrieval often reflects task-specific activation that was present during encoding. For example, a number of studies have shown that items studied alongside visual scenes, sounds, or scents during encoding differentially elicit greater levels of activation in visual association cortex, auditory association cortex, and olfactory cortices respectively during both encoding and test (e.g. Gottfried, Smith, Rugg, & Dolan, 2004; Vaidya, Zhao, Desmond, & Gabrieli, 2002; Wheeler, Petersen, & Buckner, 2000; see Danker & Anderson, 2010 for a review). More work advanced this area using multivoxel pattern analyses (MVPA), which permits examination of item-specific patterns of neural activation elicited by unique study items. In brief, a number of studies have shown that item-specific activity patterns elicited during encoding are often reinstated during successful recollection of those items, both with respect to task-relevant cortical activation (e.g. Ritchey, Wing, LaBar, & Cabeza, 2013; Wing, Ritchey & Cabeza, 2017) and activation in MTL structures (e.g. Danker, Tompary, & Davachi, 2017; Schultz et al., 2019; Staresina et al., 2012).

The present study

Participants studied a list of words, presented one at a time, in an event-related fMRI experiment. A cue indicated whether the word was to be read aloud, silently, or while saying aloud a control word ("check"). This control condition provided an estimate of baseline brain activation associated with articulating and hearing oneself produce a single word — but in a non-distinctive (i.e., non-item-specific) way. A similar manipulation was used by MacLeod et al. (2010), who showed that responding aloud with the same word ("yes" in that study) did not elicit

a production effect; hence, we did not expect our control condition to confer a memory benefit. After the study phase, participants completed a recognition test.

This design permitted a preliminary neuroimaging investigation into the mechanisms that underlie the production effect. Because distinctiveness-based accounts emphasize the additional sensorimotor processing of produced items (e.g., motor articulation), we expected stronger activation during encoding in both motor cortex (central sulcus/precentral gyrus) and auditory cortex (superior temporal gyrus) in the aloud condition than in the silent and control conditions. Presuming this information was used heuristically at test, activation in these regions was expected to correlate with the magnitude of the recollective component of the behavioural production effect. If participants differentially attended to the aloud items, stronger activation might also be expected for aloud items in areas associated with attention at encoding, such as the premotor and posterior parietal cortices (Kim, 2011). Insofar as manipulations of attention are most strongly associated with changes in recollection, activation in those regions should correlate with the magnitude of the production effect on recollection. Finally, participants might also demonstrate enhancement to other forms of encoding, such as semantic elaboration, resulting in greater activation of frontal or anterior superior temporal regions (e.g., Weber, Lau, Stillerman, & Kuperberg, 2016).

While our primary interest was in comparisons between conditions, we also examined item-specific patterns of neural activation with MVPA. One useful method of MVPA is representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008), which allows researchers to explore correspondence between item-specific neural activation and computational models that convey information about particular properties of experimental stimuli (for example, properties such as phonology, orthography, or semantic content). In the context of the current investigation, we reasoned that if produced items are indeed encoded more distinctively—owing to richer sensorimotor processing—then speaking words aloud ought to elicit more distinctive (i.e., dissimilar) activation patterns in sensorimotor regions, compared to reading them silently or saying "check" in response to every word. Moreover, activation patterns for produced items should better reflect the phonological properties of to-be-remembered words, given that phonology should approximately reflect sensory and motor information obtained through articulation. This claim would be evidenced by higher correspondence between neural data for produced items and a phonological model, compared to silent or control items.

Our primary interest was in brain activity during encoding, since the production effect is defined by how items are encoded. However, we also examined activity during the recognition test. To the extent that retrieval recapitulates encoding processes, similar regions may be recruited across study and test. In particular, if production results in relatively more elaborate memory representations, then we might see more activation in sensorimotor areas for aloud items at test. Models of the role of the parietal lobe in memory retrieval predict greater activation of the IPL for aloud items because they contain additional episodic details not present for items in the other two conditions.

Method

Subjects

Thirty-two healthy, English-speaking young adults (convenience sample; 20 females, 12 males; 20–32 years of age; $M = 24.1$ years) were recruited through on-campus advertising in exchange for $30 and an image of their brain. Previous work on the production effect has indicated that within-subjects designs typically elicit effect sizes of approximately Hedges $g = 0.6$ (Fawcett & Ozubko, 2016). A power analysis (implemented in *R* using the *pwr* package; Champely, 2020) indicated that a sample size of 26 (see below) had 80% power to detect a minimum effect size of 0.57. As such, our sample size was sufficient to detect the behavioural production effect. In the absence of prior published fMRI studies on this phenomenon, we assumed that the sample size appropriate for detecting the behavioural effect would be sufficient to identify an associated fMRI effect; as well, this number is consistent with typical sample sizes in fMRI studies published in recent years (Szucs & Ioannidis, 2020).

All participants were right-handed (Oldfield, 1971), with normal or corrected-to-normal vision, and reported no history of neurological conditions, attentional or language difficulties, current use of psychiatric medications, or contraindications to MRI scanning. The study was approved by the Dalhousie University Research Ethics Board. Participants provided informed consent according to the Declaration of Helsinki. Behavioural responses were not recorded for 4 subjects due to the response box malfunctioning; they were excluded from all behavioural and fMRI analyses. Two additional participants were excluded for appearing to confuse "know" and "no" responses at test (see Stimulus and Apparatus), or for not using each response category at least once. Thus, data from 26 participants were analyzed.

Stimuli and apparatus

Stimuli were presented using a custom script built in *PsychoPy2 1.84.2* (Peirce, 2009). The words were 120 nouns from MacDonald and MacLeod (1998), 5 letters to 10 letters in length, with frequencies greater than 30 per million (Thorndike & Lorge, 1944). For each subject, the script randomly assigned each word to one of four lists (30 words each), corresponding to the four experimental conditions (read aloud, read silently, sensorimotor control, and foils).

All words were presented in white, lowercase Courier size 20 font against a black background measuring 330 × 100 pixels superimposed in the centre of a complex visual scene that covered the remainder of the screen to reduce between-trial boredom. Study phase encoding instructions were provided using icons of a mouth (aloud condition), an eye (silent condition), or a check mark (sensorimotor control condition). These icons each measured 150 × 150 pixels, and were presented at the centre of the screen. All stimuli were presented on an LCD projector that was focused on a Mylar screen positioned in the bore behind the participants, viewed via an angled mirror. Throughout the test phase, participant responses were recorded by a fiber optic response pad (Current Designs Inc., Philadelphia, PA), using three buttons which were pressed by the index, middle, and ring fingers of one hand determined at random. Response hand was randomly determined for each subject to mitigate lateralized sensorimotor activation associated with operating the response box in the group-level contrasts. During the test phase, the mapping of the recollect, know, and no responses to these buttons was continuously presented as a reminder on-screen, centered and above the black background upon which stimulus words were presented.

Procedure

After providing informed consent and passing MRI safety screening, each participant was positioned in the MRI scanner. A brief scout scan determined head position, followed by four functional scans, and finally a structural scan. The study and test phase were both conducted in the MRI scanner. Each phase was further subdivided into two "runs" corresponding to separate fMRI scans, with a brief break in between. The first run of each phase was preceded by a practice version of the corresponding task, containing four replications of each condition. These practice phases used a unique set of words that did not appear in the main experimental fMRI runs, but were otherwise similar to their experimental counterparts.

Study phase

Participants were instructed to remember the studied items for an unspecified memory test. Each study phase trial began with a 250 ms fixation cross ("+") presented in the center of the screen to alert

participants to a new trial. The icon for that trial was then presented for 1000 ms, after which a word was presented for 2500 ms. During each study phase run, participants were presented with 15 trials of each condition (aloud, silent, sensorimotor control), presented in pseudo-random order. In addition to the 250 ms fixation period between trials, 30 additional "null" events lasting 2200 ms each were interspersed randomly to facilitate recovery of the event-related BOLD responses to each condition. These null events consisted of continuous display of the fixation cross. The placement of these null events, as well as the sequence of trials for each condition, was determined by the application *optseq2* (Dale, 1999; Dale, Greve, & Burock, 1999) to optimize recovery of the hemodynamic responses to individual stimuli (i.e., to improve estimation efficiency). The same trial order and timing were used for all participants, but the assignment of items to conditions and the order of specific items were randomized for each participant. Each study phase run lasted approximately 4 min.

Test phase

Recognition of the study items was tested using the remember-know recognition paradigm (Tulving, 1985), following a procedure detailed in Fawcett, Lawrence, and Taylor (2016). For each test item, the participant indicated whether they could *recollect* specific detail(s) about their studying of the item, *knew* they had studied the item but could not specifically recollect doing so, or *did not recognize* the item as one they had studied. Examples of each response type were provided. It was emphasized that recollect and know did not reflect differences in confidence, but rather reflected qualitative differences in how participants experienced their recognition of the items.

Each test phase trial began with a 1000 ms fixation cross ("+") presented in the center of the screen to alert participants of a new trial. A test item was then presented for 3000 ms, during which participants made their response using one of three buttons on the button box. The experiment continued after a period of 3000 ms, regardless of response.

During each test phase block, participants were presented with 15 items from each of the conditions used in the study phase (aloud, silent, sensorimotor control), as well as 15 "foil" items that were not presented during the study phase (i.e., the ratio of old items to new was 3:1). Each test phase run lasted approximately 5.5 min. As in the preceding phase, 40 null events (fixation lasting 2200 ms) were interspersed among the remaining trials according to an optimized sequence generated by *optseq2* (Dale, 1999). The same sequence of trial types and timing were used for all participants, but the order of the words was randomized for each participant.

MRI data acquisition

MRI scans were acquired on a 1.5 Tesla GE SIGNA LX MRI system (GE Medical Systems, Waukesha, WI) equipped with an 8 channel head coil. Each participant completed four functional scans followed by an anatomical scan. The fMRI scans used a gradient-echo, echo-planar pulse sequence with TR = 2 s, TE = 25 ms, flip angle = 90 deg, 64 × 64 matrix resulting in 3.75 × 3.75 mm in-plane voxel resolution with 34, 3.7 mm thick axial slices (no gap, interleaved slice acquisition). For each run, we obtained either 113 or 115 functional volumes (originally we specified 113 volumes, however after scanning 25 participants we realized that the response to the last stimulus item might be truncated, so added an additional two time points) during the study phase; 165 volumes during the test phase. Three additional volumes were acquired but automatically discarded from the start of every run immediately following acquisition. The T1-weighted anatomical image was obtained using a 3D fast spoiled gradient echo sequence (FSPGR BRAVO) with TR = 11.8 ms, TE = 4.69 ms, TI = 450 ms, flip angle = 12 deg, FOV = 202 mm, matrix = 224 × 224, 202 axial slices.

Data preprocessing and analysis

Behavioral data

Trials that did not contain a response (0.61%) were labelled as "missed" trials and were removed from analysis. The remaining behavioural data were analyzed as a function of condition (aloud, silent, control, foil) using multilevel logistic regression models (Baayen, Davidson, & Bates, 2008) implemented with the *brms* package (Bürkner, 2017a, 2017b) in R 3.5.1 (R Core Team, 2016).[1] These models were fit using a fully Bayesian approach with weakly informative priors[2]. Results are summarized on the back-transformed response (i.e., percentage) scale rather than the logit scale. Models included random intercepts and slopes for both subject and item, representing the "maximal" random structure corresponding to our design (Barr, Levy, Scheepers, & Tily, 2013).

fMRI data

The fMRI data were processed using FEAT (fMRI Expert Analysis Tool) Version 5.98, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). To mitigate potential task-related motion artefact (particularly during the aloud and sensorimotor control conditions, in which speaking aloud might cause head motion), motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002) was applied; visual inspection of the results was used to exclude any data where head motion across time points exceeded 2 mm. This resulted in the removal of 5 runs from further analysis; no more than one run per participant was removed. Additionally, 4 runs had excessive head movement only later in the runs, and so were trimmed by removing time points from the onset of excessive head motion to the end of the run (anywhere from 50 to 100 time points). Additional preprocessing steps included: non-brain removal using BET (Smith, 2002); spatial smoothing using a Gaussian kernel of FWHM 6 mm; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s). Prior to statistical analyses we excluded all trials that corresponded to an incorrect response at test from both study- and test-phase fMRI analyses.

Spatial registration and normalization was carried out using FLIRT (Jenkinson & Smith, 2001; Jenkinson et al., 2002), with each individual's EPI volumes registered to their respective high-resolution structural image (using rigid body transformation), and the high-resolution structural in turn registered to the MNI152 template using first linear affine, and then nonlinear methods (the latter implemented in FNIRT; Andersson et al., 2007a, 2007a). The outputs of first-level statistical analysis (see next paragraph) were transformed to standard space using the combined EPI-to-structural and structural-to-MNI152 transforms, and resliced to 2 mm isotropic resolution.

Statistical analyses of the fMRI data proceeded over three levels. The

[1] Each model was fit using 10,000 iterations with 5000 warm-up samples; convergence was verified through visual inspection and using standard convergence metrics such as R-hat ≈ 1 (Gelman & Hill, 2006). There were no divergent transitions. For more detail see Fawcett and Ozubko (2016) and Fawcett et al. (2016).

[2] For the "old" and independence know data, priors for the intercept (false alarm rate) and slopes were represented by *Normal*(-1, 2) and *Normal*(0, 4), respectively. Priors for the intercept were broad but acknowledged that false alarms were likely to be rare (i.e., < 50%); priors for the slopes were effectively uniform. For the recollect data, priors for the intercept (false alarm rate) and slopes were represented by *Normal*(-4, 2) and *Normal*(0, 4), respectively. These changes reflected our knowledge that false alarms would be less common for recollect responses. Priors for the SD of each random effect were represented by *Normal*(0, 2) with a regularizing prior on the correlation matrix equivalent to *LKJ*(4). All priors are reported on the logit scale. Models fit instead using (less principled) default priors provided by the *brms* package produced similar results.

first level was performed on each run individually, and involved multiple linear regression using FSL's FILM with local autocorrelation correction (Woolrich, Ripley, Brady, & Smith, 2001). Regressors included time series for each stimulus type (aloud, silent, sensorimotor control; foils for test phase runs) convolved with a model of the hemodynamic response (a gamma function), as well as the six parameters derived from the motion correction step as covariates of no interest. The regressors of interest were orthogonalized with respect to the motion parameters to eliminate issues of collinearity. Contrasts of interest included each condition relative to baseline, as well as the pairwise contrasts aloud-silent, aloud-control, and control-silent. For the test phase, we also contrasted each "old" condition (aloud, control, silent) against the foil items.

The parameter estimates and associated variances from first-level analyses were combined in the second-level analysis, separately for each participant and phase (study/test) using fixed-effects linear regression to estimate the mean effect across runs for each contrast for each participant.

Finally, the third-level analysis was performed using nonparametric permutation inference with FSL's *randomise* (Winkler, Ridgway, Webster, Smith, & Nichols, 2014). Correction for multiple comparisons of the resulting statistical maps was applied using threshold-free cluster enhancement (Smith & Nichols, 2009) with family-wise error correction set at $p < .05$. For between-condition contrasts, the data were masked during the nonparametric inference procedure to restrict the analyses to voxels that were significantly ($p \leq 0.05$) activated in the nonparametric analysis of the minuend for that contrast relative to baseline. For example, the aloud-control contrast was restricted to voxels that were significant in the aloud-baseline contrast. Tables of the resulting activations were generated using FSL's *cluster* routine to identify clusters of contiguous voxels (with a minimum spatial extent of 25 adjacent voxels, to exclude small clusters that were likely spillover from another ROI) and the location of the peak *z* score within each cluster, and *atlasquery* to identify the anatomical label of the voxel having the peak *z* value within each cluster.

We conducted additional analyses in which we correlated fMRI data from the study phase either with test-phase performance in each condition corrected for false alarms[3], or with the behavioural production effect[4]. Both behavioural measures were calculated three ways: (1) overall 'old' performance (percentage of items correctly identified as old at test), (2) 'independent know' performance (percentage of old items to which participants made a 'know' response after excluding trials with a 'recollect' response), and (3) recollection performance (percentage of old items to which participants made a 'recollect' response). This resulted in six behavioral scores for each participant: three accuracy scores and three production effect scores, computed separately for old, know, and recollect performance.

We correlated the BOLD response for each condition during the study phase relative to baseline with the six behavioural scores. Here, the first- and second-level analyses included *all* trials from the study phase (i.e., we did not exclude items that were incorrectly identified as 'new' at test). We then conducted three sets of third-level analyses in which the BOLD response for each condition (aloud, silent, control) was correlated with behavioural scores: either overall performance in the corresponding condition (e.g., BOLD responses to aloud trials were correlated with accuracy for aloud items), or the behavioural production effect. This was achieved by conducting third-level analyses as described earlier, but in this case the behavioural score of interest was included as a covariate. Finally, the results from these correlation analyses were masked with activation maps from our main analyses (described above) to ensure that correlations with behaviour were restricted to areas showing task-related activation. (see Table 1)

This procedure was repeated for each condition, resulting in 18 separate correlation analyses, each including one behavioural score as a single covariate. Finally, we replicated this procedure to correlate the BOLD response derived from the aloud-silent and aloud-control contrasts during the study phase with the behavioural production effect (aloud-silent and aloud-control respectively) separately for old, know, and recollect judgments. Having conducted these 24 correlation analyses, we corrected for multiple comparisons by applying a Bonferroni adjustment (i.e., alpha level / 24) to the results of any correlation that yielded significant results.

Most of the fMRI analyses resulted in very large clusters spanning multiple brain regions; in many cases, all activations were subsumed in a single contiguous cluster. To produce tables that accurately represented the brain regions included in these large activation clusters, we performed clustering for each activation map of interest, within each brain region defined in the Oxford-Harvard cortical and subcortical atlases (e.g., Newman et al., 2010a, 2010b, 2015). In all cases, the statistical analyses for each contrast were performed on the entire brain and corrected using threshold-free cluster enhancement; this segmentation into regions of interest (ROIs) was performed only for the purpose of generating Tables 2–6.

With respect to results from the test phase, because we were primarily interested in activation that reflected encoding processes (i.e., reinstatement), activation derived from each contrast at test was spatially constrained to areas that were activated for the same contrast at study (e.g., results from the aloud-baseline contrast at test were masked with the aloud-baseline contrast at study, and so on). Contrasts involving foils were constrained to the contrast of the minuend condition relative to baseline at study (e.g., the aloud–foil contrast was masked with aloud–baseline from the study phase, and so on). The only test phase contrast that was not spatially constrained in this manner was foil–baseline. Results for non-masked test phase contrasts are reported in supplementary material.

Representational similarity analysis (RSA)

Procedures for our multivariate analyses are detailed in supplementary materials, so we will describe them only briefly here. We first obtained single-trial estimates of activation for every item presented during the study phase using an iterative modelling procedure proposed by Mumford, Turner, Ashby, and Poldrack (2012). Mathematically, single-trial estimates provide a pattern vector for every trial, whereby each value in the vector indicates the level of activation in a particular voxel. We performed RSA to assess correspondence between these neural pattern vectors and a formal phonological model, using a whole brain searchlight analysis. More specifically, at the center of every searchlight sphere (3 mm radius) we constructed a neural dissimilarity matrix (DSM) comprising pairwise correlation distances between the activation patterns to each item, for each condition and subject separately. These neural DSMs were correlated with a phonological model: a DSM comprising pairwise phonological edit distances reflecting phonological dissimilarity between study items. Our searchlight analysis used functions from the CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016) implemented in MATLAB (The MathWorks, Inc., Natick) and additional custom code.

To avoid biasing the MVPA results with the results from our univariate analysis, the results from our searchlight analysis were constrained to a set of independent, a priori ROIs relevant to word reading identified by Murphy, Jogia, and Talcott (2019; see Table S1 for the list of ROIs and their MNI coordinates). Within these ROIs, we used random-

[3] This correction involved subtracting the percentage of false alarms (i.e. "remember" or "know" responses to foil items) from the percentage of hits to items that were present during the study phase. This correction was intended to control for participants guessing at test.

[4] For correlations involving neural data from the aloud or silent conditions, the production effect was defined as performance on aloud trials minus performance on silent trials. For correlations involving neural data from the control condition, the production effect was defined as performance on aloud trials minus performance on control trials.

Table 1
Activation coordinates for each condition relative to baseline during the study phase.

Study Phase			Aloud-Baseline					Silent-Baseline					Control-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	943	5.43	0	18	36						357	4.25	−2	6	44
		RH	518	5.23	2	18	34						39	3.74	2	4	46
	Frontal Operculum Cortex	LH	426	5.42	−50	10	−2						388	5.04	−50	10	−2
		RH	264	4.31	38	12	8						278	4.28	50	10	−2
	Frontal Orbital Cortex	LH	507	5.07	−38	28	−6						702	4.84	−38	32	−6
			103	4.32	−22	4	−14										
		RH	122	4.18	24	6	−10						35	2.98	34	26	0
			90	3.44	38	28	−2										
	Frontal Pole	LH	96	3.9	−48	38	6	161	2.96	−40	38	10	2540	4.82	−32	48	28
		RH	152	3.18	50	38	10										
	Inferior Frontal Gyrus pars opercularis	LH	551	5.31	−54	10	−2	387	3.73	−44	14	18	1189	5.76	−60	14	−4
		RH	254	3.77	48	8	14						98	3.66	40	20	12
			46	3.59	60	12	0						93	3.79	52	14	−2
	Inferior Frontal Gyrus pars triangularis	LH	327	4.09	−44	30	−2	390	3.89	−46	28	22	991	4.58	−44	24	20
		RH	74	2.92	40	32	4						77	3.6	56	20	−8
													30	3.4	42	20	14
	Middle Frontal Gyrus	LH	237	5.67	−40	−2	58	611	4.21	−46	28	24	2585	5.5	−32	−2	58
								535	4.14	−38	−2	60					
		RH	139	4.8	56	6	48						356	4.57	40	0	62
	Paracingulate Gyrus	LH	359	5.01	−2	18	38	435	3.83	−6	10	52	646	5.61	0	10	52
		RH	125	4.52	2	18	38	29	2.88	2	8	50	126	5.3	2	10	54
	Precentral Gyrus	LH	4447	7.03	−44	−14	36	1163	4.44	−48	2	50	4308	6.11	−50	−6	48
		RH	3536	7.08	46	−10	34						3140	6.45	54	−2	44
	Subcallosal Cortex	LH	30	3.88	0	6	2										
	Superior Frontal Gyrus	LH	948	4.89	0	10	60	395	3.7	−4	10	56	1798	5.63	−4	10	60
		RH	645	5.62	6	10	62						626	5.28	2	10	56
	Supplementary Motor Area	LH	927	5.9	0	4	62	338	4.72	−6	6	54	889	6.44	−2	2	62
		RH	735	6.51	2	4	62	66	3.42	2	6	54	747	6.21	2	4	60
Parietal	Angular Gyrus	LH						187	3.52	−44	−52	42	409	5.09	−42	−58	54
													158	4.56	−62	−50	12
	Central Opercular Cortex	LH	1084	7.08	−58	−10	6						911	5.97	−62	−10	6
		RH	840	6.84	60	−4	10						762	6.04	60	−8	12
	Cingulate Gyrus posterior	LH	102	4.33	−2	−44	0						155	3.76	−4	−48	−4
		RH	80	4.08	4	−42	0						58	3.37	2	−44	2
	Parietal Operculum Cortex	LH	360	4.6	−62	−26	16						153	4.24	−64	−38	20
		RH	145	4.47	64	−28	20						44	3.05	56	−30	18
	Postcentral Gyrus	LH	2496	7.1	−46	−14	30						2330	6.15	−50	−12	24
			198	3.48	−2	−50	74										
		RH	1776	7.43	46	−12	34						1506	6.13	48	−12	30
	Precuneous Cortex	LH	216	3.92	−2	−80	52						689	5.03	−2	−82	48
		RH	74	3.56	2	−68	64						474	4.75	2	−58	72
			49	3.38	2	−82	50										
	Superior Parietal Lobule	LH	32	3.06	−22	−44	48						489	4.61	−38	−56	52
	Supramarginal Gyrus anterior	LH	116	4.33	−64	−24	16						95	3.48	−52	−40	46
													41	3.82	−64	−24	16
		RH	116	4.24	72	−16	12						78	4.36	72	−16	12
	Supramarginal Gyrus posterior	LH	391	4.73	−52	−44	10	42	3.35	−42	−52	42	1193	5.59	−64	−44	10
		RH	264	4.91	52	−38	6						304	5.58	52	−38	6
Temporal	Heschls Gyrus	LH	343	6.77	−56	−10	4						359	5.2	−56	−18	8
		RH	344	6.47	54	−14	2						295	5.17	56	−12	6
	Inferior Temporal Gyrus posterior	LH	111	4.44	−46	−44	−16						299	4.27	−48	−46	−8

(*continued on next page*)

Table 1 (*continued*)

Study Phase			Aloud-Baseline					Silent-Baseline					Control-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
	Inferior Temporal Gyrus temporooccipital	LH	764	5.55	−48	−54	−18	394	4.41	−46	−64	−16	1053	6.35	−48	−54	−24
		RH	616	6.04	56	−62	−20	199	4.32	46	−60	−16	789	5.38	48	−60	−16
	Lingual Gyrus	LH	1744	6.7	−14	−64	−18	215	5.95	−16	−88	−6	1801	6.08	−16	−88	−6
		RH	1519	6.95	20	−64	−20	159	6.14	16	−86	−8	1488	6.85	14	−88	−8
	Middle Temporal Gyrus anterior	LH	240	4.8	−62	2	−12						125	4.25	−62	2	−12
		RH	128	5.71	62	−6	−10						31	3.67	64	−4	−10
	Middle Temporal Gyrus posterior	LH	416	4.77	−68	−20	−6						964	5.09	−66	−42	4
		RH	680	5.21	72	−28	0						764	5.48	52	−34	−2
	Middle Temporal Gyrus temporooccipital	LH	217	4.58	−52	−44	6						957	5.55	−58	−58	8
		RH	160	5.22	52	−38	4						635	5.63	46	−38	4
	Planum Polare	LH	494	7.03	−56	0	−4						244	5.31	−52	4	−4
		RH	349	6.53	58	−2	−2						163	4.99	58	−6	4
	Planum Temporale	LH	765	7.2	−58	−10	4						729	5.83	−62	−10	4
		RH	552	7.07	62	−14	2						543	5.48	64	−8	6
	Superior Temporal Gyrus anterior	LH	404	7.05	−56	−2	−4						375	5.69	−64	−10	4
		RH	388	6.81	60	−6	−4						309	4.84	60	2	−2
	Superior Temporal Gyrus posterior	LH	1348	7.07	−62	−20	2						1331	5.84	−70	−10	4
		RH	1418	7.24	62	−14	0						1371	5.6	66	−18	0
	Temporal Fusiform Cortex anterior	LH	236	3.87	−32	−2	−34										
		RH	114	3.21	24	−2	−44										
	Temporal Fusiform Cortex posterior	LH	575	4.94	−34	−52	−26	73	3.55	−38	−42	−24	596	6.41	−40	−50	−28
		RH	93	3.6	36	−42	−30						103	4.56	36	−42	−30
			29	2.83	38	−8	−24						31	3.23	38	−16	−18
	Temporal Occipital Fusiform Cortex	LH	719	5.84	−18	−64	−18	488	4.94	−38	−62	−16	729	6.56	−44	−52	−28
		RH	735	6.58	36	−62	−20	547	5.52	42	−58	−18	768	5.74	42	−60	−26
	Temporal Pole	LH	1147	6.78	−56	4	−6						508	6	−52	8	−6
			77	3.49	−34	2	−36										
		RH	1040	6.18	54	8	−12						433	5.16	54	10	−6
			245	4.23	26	0	−18										
Occipital	Cuneal Cortex	LH	302	4.53	−6	−86	40						333	4.86	0	−88	44
		RH	384	4.66	2	−86	36						386	5.26	4	−86	34
	Intracalcarine Cortex	LH	672	5.86	−4	−72	8	81	4.46	−8	−90	−2	712	4.92	−18	−86	10
		RH	913	5.1	2	−74	10	30	3.57	12	−88	0	740	5	12	−88	0
	Lateral Occipital Cortex inferior	LH	2345	6.36	−40	−84	−10	1700	5.95	−36	−88	−4	2258	6.05	−24	−90	4
			35	2.96	−52	−62	8						60	4.4	−60	−62	6
		RH	2628	6.65	44	−78	−16	1970	6.92	30	−86	−6	2277	6.14	44	−78	−16
	Lateral Occipital Cortex superior	LH	454	5.22	−30	−84	8	802	3.78	−32	−64	40	2664	5.29	−40	−58	52
								195	4.92	−20	−90	8					
		RH	512	4.87	40	−88	8	796	4.96	26	−88	6	562	5.01	34	−86	16
													266	3.84	6	−74	60
	Occipital Fusiform Gyrus	LH	1515	6.73	−24	−82	−12	1371	6.61	−22	−82	−12	1652	6.63	−34	−74	−20
		RH	1481	6.89	16	−68	−20	1159	6.49	30	−84	−8	1495	6.61	26	−86	−12
	Occipital Pole	LH	2999	6.63	−14	−94	−10	2330	6.3	−16	−90	−8	2783	6.47	−20	−94	2
		RH	2678	6.28	18	−94	−14	1644	5.99	24	−90	−6	2360	6.45	14	−90	−8
	Supracalcarine Cortex	LH	80	5.07	0	−74	12						78	3.87	0	−74	12
		RH	91	4.94	2	−74	12						88	3.97	2	−74	12
Medial	Amygdala	LH	327	5.04	−20	−2	−18						128	3.26	−16	−4	−20
		RH	353	4.45	26	−2	−16						26	3.21	28	−2	−14
	Hippocampus	LH	474	5.26	−20	−26	−8						361	3.73	−22	−26	−10
		RH	297	3.38	22	−12	−22						140	3.51	34	−14	−18
	Insular Cortex	LH	1201	4.94	−38	2	2						930	4.61	−34	20	−2

(*continued on next page*)

Table 1 (*continued*)

Study Phase			Aloud-Baseline					Silent-Baseline					Control-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
		RH	1220	5.4	36	8	2						835	4.35	32	6	2
	Parahippocampal Gyrus anterior	LH	610	5.04	−20	−2	−18						106	3.26	−16	−4	−20
													52	3.73	−36	−20	−18
		RH	586	4.22	28	0	−18						71	3.51	34	−14	−18
	Parahippocampal Gyrus posterior	LH	469	5.26	−20	−26	−8						252	4.35	−2	−44	−4
			27	3.22	−18	−36	−24										
		RH	143	3.67	10	−30	−8										
Basal Ganglia	Caudate	LH	389	5.5	−10	0	10						230	3.85	−10	−2	12
		RH	231	4.5	8	2	8						37	3.04	18	4	18
	Pallidum	LH	243	4.62	−18	4	2						139	4.21	−20	2	0
		RH	234	4.78	20	4	2						168	3.88	26	−10	−2
	Putamen	LH	740	5.08	−26	6	4						664	4.57	−22	4	0
		RH	696	5.05	22	6	6						624	4.41	26	6	2
Midbrain	Thalamus	LH	1065	5.75	−10	−2	8						699	4.21	−10	−4	10
		RH	941	5.09	6	−2	8						549	4.28	6	−2	4

effect cluster statistics to identify any clusters which significantly differed between conditions.

Results

Behavioural

Recollect and know responses during the recognition test were initially aggregated into "old" responses and analyzed as a function of item type (aloud, silent, control, foil). Mean proportions corresponding to each condition are plotted in Fig. 1. Contrasts for differences between the individual conditions were calculated using the *emmeans* package (Lenth, 2018). Participants were more likely to respond "old" to silent items than to foil items, *difference* = 43.4%, *95% HDI* [37.0%, 50.0%], but there was little difference in the recognition of silent and control items, *difference* = 1.9%, *95% HDI* [− 5.7%, 9.8%]. The latter finding supports previous work showing that producing a non-unique response to items does not reliably improve memory (MacLeod et al., 2010). Importantly, we also replicated the standard production effect: Recognition was greater for aloud items compared to either silent items, *difference* = 21.3%, *95% HDI* [13.9%, 28.5%], or control items, *difference* = 19.4%, *95% HDI* [10.6%, 27.9%].

We next evaluated the effect of production on recollection and familiarity. For familiarity, we first excluded all trials for which a recollect response had been made, to correct for nonindependence of remember/know judgments that can underestimate familiarity (e.g., Jacoby, Yonelinas, & Jennings, 1997; Mangels, Picton, & Craik, 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas, 2002; Yonelinas & Jacoby, 1995). In the context of a logistic regression model, fitting the model for "know" responses after excluding "recollect" responses produces an estimate comparable to standard independent remember/know calculations (for statistical proof, see Fawcett et al., 2016; Fawcett & Ozubko, 2016).

Complementary models were fit to the proportion of recollect and independent know responses made for each item type. Separate models were fit to each judgment because recollect and know judgments were mutually exclusive and hence dependent. Mean proportions are also reported in Fig. 1. For the recollection model, participants were more likely to recollect silent items than foil items, *difference* = 11.6%, *95% HDI* [6.8%, 16.9%], whereas recollection of the silent items again failed to differ from the recollection of the control items, *difference* = 2.3%, *95% HDI* [−3.0%, 7.6%]. Recollection was greater for aloud items compared to either silent items, *difference* = 14.1%, *95% HDI* [7.7%, 21.3%], or control items, *difference* = 11.9%, *95% HDI* [5.0%, 19.2%].

For the familiarity model, silent items were more familiar than foil items, *difference* = 35.2%, *95% HDI* [27.8%, 42.7%], but not control items, *difference* = 0.0%, *95% HDI* [9.4%, 9.2%]. Familiarity was again greater for aloud items compared to either silent items, *difference* = 20.9%, *95% HDI* [12.3%, 29.7%], or control items, *difference* = 20.9%, *95% HDI* [10.2%, 31.4%]. These findings replicate the production effects on recollection and familiarity in past behavioural studies (Fawcett & Ozubko, 2016; Ozubko et al., 2012).

We calculated the Cohen's *d* effect size of the behavioural production effect, defined as the standardised difference between produced and silent items, for each measure separately (using the R package *rstatix*; Kassambara, 2021). This yielded effect sizes of $d = 1.36$, 1.00, and 1.29 for the combined "old", recollection, and familiarity judgments respectively. Finally, we assessed the split-half reliability for the observed production effect using a permutation-based approach (implemented in *R* using the *splithalf* package; Parsons, 2020) with 5000 random splits. In brief, this entailed iteratively splitting the data from each participant into two halves (without replacement) such that, for each permutation, the production effect was calculated twice. Production effect scores were

Table 2
Activation coordinates for contrasts between conditions during the study phase.

Study Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	787	4.64	−2	16	32						243	4.04	−8	2	42
		RH	210	4.2	2	16	32										
			65	3.85	2	−8	46										
	Frontal Operculum Cortex	LH											155	3.16	−48	16	−6
		RH	52	3.49	52	10	0										
	Frontal Orbital Cortex	LH	47	3.5	−20	4	−16						170	3.89	−54	20	−8
		RH	77	3.36	22	4	−16										
	Inferior Frontal Gyrus pars opercularis	LH	75	4.39	−60	12	−4						298	4.78	−60	12	−4
		RH	45	3.35	58	10	14										
	Inferior Frontal Gyrus pars triangularis	LH											54	3.63	−52	20	−8
	Middle Frontal Gyrus	LH											383	3.94	−38	0	56
	Paracingulate Gyrus	LH											42	3.52	0	10	54
	Precentral Gyrus	LH	2520	6.77	−48	−10	30						1640	6.22	−52	−6	24
													312	4.7	0	−14	68
		RH	2350	6.9	50	−8	26	65	2.9	60	2	18	1825	5.97	48	−8	30
			121	4.59	2	−14	68						206	4.35	2	−14	68
	Superior Frontal Gyrus	LH	356	4.12	−16	−2	66						399	4.31	−8	−6	72
													80	3.31	−24	4	46
		RH	278	4.18	6	10	64						191	3.96	8	6	68
	Supplementary Motor Area	LH	561	4.74	0	−10	66						646	5.31	−2	−8	68
		RH	537	4.76	2	2	62						466	5.17	2	2	64
Parietal	Angular Gyrus	LH											114	4.09	−56	−60	14
													44	3.77	−40	−54	50
	Central Opercular Cortex	LH	921	6.66	−58	−10	6	124	3.5	−60	−10	6	733	6.7	−62	−10	6
		RH	670	6.48	62	−4	12	62	3.18	62	−6	6	684	6.07	58	−4	12
	Cingulate Gyrus posterior	LH	82	4.4	−2	−44	0						119	3.75	−4	−48	−4
		RH	65	4.59	4	−38	0						38	3.34	2	−44	2
	Parietal Operculum Cortex	LH	302	5.41	−62	−26	16						142	4.75	−62	−24	16
		RH	111	4.11	62	−26	18						37	3.44	64	−24	18
	Postcentral Gyrus	LH	2082	6.78	−44	−12	28	57	3.36	−44	−14	30	1334	6.72	−52	−10	24
													243	3.34	−4	−44	70
													27	2.61	−44	−22	62
		RH	1151	7.05	52	−10	30	75	3.1	66	−10	12	927	6.2	50	−10	30
			204	4.19	26	−28	70						105	3.26	4	−40	72
													87	3.34	24	−30	60
	Precuneous Cortex	LH											550	3.98	0	−78	36
		RH											233	3.67	2	−78	38
	Superior Parietal Lobule	LH											150	3.99	−38	−54	52
	Supramarginal Gyrus anterior	LH	116	5.18	−64	−24	16						41	5.21	−64	−24	16
		RH	93	3.72	64	−20	22						75	4.02	66	−22	18
	Supramarginal Gyrus posterior	LH	207	4.22	−46	−42	12						481	4.6	−62	−42	14
		RH	177	4.35	60	−36	12						207	5.2	48	−38	8
Temporal	Heschls Gyrus	LH	280	6.55	−56	−10	4	52	4.25	−56	−12	2	317	5.06	−56	−10	4
		RH	296	6.2	50	−18	4	80	4.24	58	−6	0	272	5.9	54	−14	8
	Inferior Temporal Gyrus posterior	LH											48	3.98	−48	−42	−8
	Inferior Temporal Gyrus temporooccipital	LH	125	3.1	−46	−46	−14						231	3.3	−52	−62	−14
													37	3.57	−46	−46	−8
		RH	91	3.06	54	−52	−22						159	3.32	64	−50	−20
	Lingual Gyrus	LH	1110	5.61	−16	−64	−18						1185	5.25	−16	−64	−18
		RH	1043	6.3	16	−64	−18						1025	5.77	16	−64	−18

(*continued on next page*)

Table 2 (*continued*)

Study Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
	Middle Temporal Gyrus anterior	LH	223	5.17	−64	−8	−8						87	4.21	−68	−10	−6
		RH	128	4.63	60	4	−16	69	3.3	66	−2	−10	31	3.66	62	2	−14
	Middle Temporal Gyrus posterior	LH	204	4.61	−64	−12	−8						617	4.53	−66	−26	−4
		RH	635	5.83	54	−20	−8						673	6.19	56	−20	−6
	Middle Temporal Gyrus temporooccipital	LH	29	3.11	−48	−44	6						510	4.46	−56	−58	8
													31	2.85	−60	−60	−8
		RH	97	3.98	50	−38	4						437	4.99	46	−40	6
	Planum Polare	LH	460	6.66	−60	−8	4	53	4.05	−56	0	−4	230	5.56	−60	−8	4
		RH	315	6.27	58	−2	−2	77	4.3	62	−4	2	162	4.88	66	−4	6
	Planum Temporale	LH	752	7.17	−58	−10	4	159	5.23	−62	−14	2	705	6.82	−64	−10	6
		RH	471	6.21	62	−10	0	156	4.68	62	−8	0	484	5.95	56	−14	6
	Superior Temporal Gyrus anterior	LH	401	6.93	−60	−10	2	205	5.14	−60	−12	0	374	6.64	−64	−10	4
		RH	387	6.14	58	−2	−4	299	4.56	62	−6	0	309	4.84	62	−6	0
	Superior Temporal Gyrus posterior	LH	1254	6.93	−70	−10	4	323	5.46	−64	−14	0	1292	6.77	−66	−10	4
		RH	1347	6.54	58	−28	4	375	4.52	64	−8	0	1321	6.17	56	−18	−6
	Temporal Fusiform Cortex anterior	LH	82	3.28	−40	−6	−28										
	Temporal Fusiform Cortex posterior	LH	161	3.7	−30	−34	−32						118	3.87	−46	−44	−8
	Temporal Occipital Fusiform Cortex	LH	211	5.42	−22	−64	−20						99	5.13	−18	−64	−18
		RH	120	4.84	20	−60	−18						64	4.98	22	−62	−18
													38	3.29	48	−52	−30
	Temporal Pole	LH	827	5.54	−54	6	−10						446	5.21	−56	14	−8
			33	2.97	−28	2	−32										
		RH	880	5.9	54	6	−8	248	3.81	58	6	−8	378	4.66	54	10	−6
			152	3.69	34	4	−26										
Occipital	Cuneal Cortex	LH	167	3.8	−2	−76	22						279	4.2	−4	−84	32
		RH	161	3.68	2	−74	22						223	3.86	2	−78	36
	Intracalcarine Cortex	LH	365	4.14	−6	−76	10						353	3.15	−8	−78	10
		RH	533	4.36	24	−64	8						397	4.1	18	−62	6
	Lateral Occipital Cortex inferior	LH	215	3.47	−42	−84	−24						194	3.72	−56	−66	−10
													57	3.97	−56	−64	12
		RH	52	3.08	60	−68	−10										
	Lateral Occipital Cortex superior	LH											516	3.95	−12	−62	56
	Occipital Fusiform Gyrus	LH	204	5.34	−16	−68	−20						225	5.07	−16	−68	−20
			29	3.43	−10	−92	−24										
			27	2.65	−46	−68	−24										
		RH	270	6.19	16	−68	−20						328	5.19	22	−62	−20
	Occipital Pole	LH	41	3.28	−8	−94	−24						144	3.6	−2	−92	0
													59	3.56	0	−94	−16
		RH	35	3.14	10	−92	−24						271	2.96	8	−92	6
													76	4.22	6	−94	−20
	Supracalcarine Cortex	LH	53	3.55	0	−74	20						44	2.99	0	−74	20
		RH	45	3.15	2	−74	18						26	2.9	2	−78	10
Medial	Amygdala	LH	296	4.1	−22	−4	−12						55	3.16	−20	−14	−12
		RH	261	4.28	22	−2	−14										
	Hippocampus	LH	226	3.97	−18	−24	−12						116	3.6	−20	−24	−10
	Insular Cortex	LH	634	4.66	−36	−10	12						149	3.91	−36	−10	10
													51	3.04	−42	6	0
		RH	630	4.02	36	8	2						26	2.81	32	14	8
													25	3.38	46	8	−4
	Parahippocampal Gyrus anterior	LH	303	3.76	−20	2	−16										
		RH	142	4.12	22	−2	−16										
	Parahippocampal Gyrus posterior	LH	251	4.57	−2	−44	−2						81	3.45	−20	−24	−12

(*continued on next page*)

Table 2 (*continued*)

Study Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
													44	3.28	−36	−32	−12
													26	4.45	−2	−44	−4
		RH	69	3.88	8	−36	−2										
Basal Ganglia	Caudate	LH	60	3.35	−14	12	10										
			25	3.45	−8	0	10										
	Pallidum	LH	194	3.79	−26	−16	−4						51	3.19	−24	−4	−4
		RH	152	4.05	22	−2	−4						84	3.66	20	−2	−2
	Putamen	LH	536	4.22	−30	−16	−4						139	3.4	−30	−10	−6
		RH	466	4.01	22	8	6						229	3.21	26	8	−6
Midbrain	Thalamus	LH	722	4.44	−8	−4	8						302	3.63	−14	−16	0
		RH	771	4.2	10	−4	8						381	3.96	8	−2	2

Table 3
Activation coordinates for each condition relative to baseline during the test phase, constrained to areas that were active for the same contrasts during the study phase (* but note that the Foil-baseline contrast was not constrained to study phase activation).

Conjunction Study & Test Phase			Aloud-baseline					Silent-baseline					Control-baseline					Foil-baseline *				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	504	6.07	−4	16	38						346	5.46	−4	16	38	130	4.23	−4	12	40
		RH	175	5.18	4	20	36						37	4.44	2	18	36	103	3.45	6	8	44
	Frontal Operculum Cortex	LH	332	5.16	−46	16	0						296	5.55	−50	12	−4					
		RH	63	4.24	38	24	0						45	4.12	48	16	2					
	Frontal Orbital Cortex	LH	464	5.68	−32	26	−2						611	5.12	−32	24	−4	54	3.08	−46	16	−8
			41	3.17	−20	10	−10															
		RH	85	4.72	34	28	−4															
			49	4.25	18	4	−12															
	Frontal Pole	LH	89	3.96	−46	40	8	27	3.37	−42	44	8	1882	5.01	−42	48	8					
	Inferior Frontal Gyrus pars opercularis	LH	551	5.72	−60	12	20	220	3.95	−36	18	22	1143	5.83	−50	14	−4	366	3.86	−50	10	28
		RH	95	3.75	60	12	16						48	3.93	50	16	0					
			27	3.6	52	16	−2															
	Inferior Frontal Gyrus pars triangularis	LH	245	4.31	−50	18	−6	264	4.66	−50	28	22	923	4.73	−50	26	24	126	3.33	−46	28	18

(*continued on next page*)

Table 3 (*continued*)

Conjunction Study & Test Phase			Aloud-baseline					Silent-baseline					Control-baseline					Foil-baseline *				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
		RH											46	3.94	52	18	−4					
	Middle Frontal Gyrus	LH	210	5.87	−34	−4	58	546	4.89	−46	26	26	2157	5.38	−34	−4	58	431	4.77	−32	−4	56
								357	5.03	−36	−4	58						76	3.09	−46	24	24
		RH	42	4.21	34	−2	62						206	4.7	34	−2	50	60	3.19	36	−2	60
	Paracingulate Gyrus	LH	342	6.69	−2	14	52	379	5.69	−2	18	40	589	6.57	−4	20	42	379	4.91	−2	14	48
		RH	116	5.91	2	12	52	29	4.36	2	18	42	103	5.29	2	20	38	316	4.52	2	10	48
	Precentral Gyrus	LH	2697	6.41	−50	2	34	784	5.52	−36	−8	58	2439	5.41	−46	4	42	2862	5.24	−52	−2	36
			230	3.96	−6	−14	70						139	4.01	−6	−14	70					
		RH	586	4.75	58	6	28						501	4.72	30	−4	48	305	3.56	40	−8	60
			311	4.77	38	−10	60															
	Superior Frontal Gyrus	LH	805	6.39	−2	14	54	223	4.32	−6	22	48	892	5.45	−4	24	48	165	3.64	−28	−4	62
																		74	4.24	−2	16	52
																		48	3.15	−8	−6	68
		RH	239	5.65	2	14	54						78	4.33	2	18	52	162	3.38	10	−4	70
			34	4.04	30	−4	62						31	3.49	30	−4	60					
	Supplementary Motor Area	LH	804	5.84	−2	2	62	246	4.25	−4	−2	66	706	4.72	−2	−2	64	503	4.36	0	6	50
		RH	494	5.45	2	6	58	45	3.25	4	6	50	320	3.95	2	8	52	379	4.42	2	6	50
Parietal	Angular Gyrus	LH						156	5.04	−38	−56	38	334	5.85	−36	−56	36					
	Central Opercular Cortex	LH	282	4.57	−46	2	14						346	4.6	−50	8	−2	214	4.33	−54	−20	20
			90	4.36	−56	−22	20											28	2.6	−40	−6	12
	Cingulate Gyrus posterior	LH	79	4.03	0	−40	2						47	3.34	0	−32	6					
		RH	29	3.57	2	−40	2															
	Parietal Operculum Cortex	LH											48	3.95	−48	−28	14	240	4.28	−50	−24	18
	Postcentral Gyrus	LH	880	5.03	−46	−20	52						774	5.3	−48	−24	36	2304	4.6	−38	−28	64
			184	3.92	−2	−40	70						199	3.74	−6	−46	66					
		RH	105	4.17	56	−14	28						107	3.95	52	−16	40	61	3.33	42	−36	52
			28	3.24	2	−38	70						49	3.93	42	−18	50					
	Precuneous Cortex	LH	187	3.94	−2	−82	56						563	5.16	−6	−72	48					
		RH											59	3.74	2	−76	42					
	Superior Parietal Lobule	LH											486	6.4	−42	−44	46	1038	4.79	−28	−58	48
		RH																377	3.75	42	−38	52
	Supramarginal Gyrus anterior	LH											94	6.44	−46	−40	46	839	4.74	−48	−36	46
		RH																55	3.11	46	−32	46
	Supramarginal Gyrus posterior	LH	83	3.57	−46	−42	30	41	5.04	−46	−48	46	538	6.52	−44	−42	46	226	4.26	−46	−40	48
													45	3.3	−48	−50	12	81	3.63	−56	−42	18
		RH																106	3.82	44	−38	52
Temporal	Inferior Temporal Gyrus posterior	LH	71	4.8	−46	−44	−20						113	4.39	−46	−44	−20	25	3.03	−46	−42	−20
	Inferior Temporal Gyrus temporooccipital	LH	724	6.28	−46	−62	−18	340	4.66	−50	−62	−24	646	6.21	−42	−62	−10	437	5.88	−46	−62	−18
		RH	187	4.62	48	−62	−20						129	4.61	46	−60	−16	225	4.31	46	−60	−16
	Lingual Gyrus	LH	750	5.5	−12	−90	−12	176	5.18	−12	−90	−12	389	5.18	−12	−90	−8	472	5.74	−12	−90	−12
		RH	760	6.85	8	−78	−26	142	5.9	14	−88	−6	497	6.45	16	−88	−8	345	5.82	16	−88	−8

(*continued on next page*)

Table 3 (*continued*)

Conjunction Study & Test Phase			Aloud-baseline					Silent-baseline					Control-baseline					Foil-baseline *				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
	Middle Temporal Gyrus posterior	LH	154	4.17	−66	−34	−2						430	3.82	−68	−38	−4					
	Middle Temporal Gyrus temporooccipital	LH	170	4.72	−56	−52	6						471	3.66	−56	−54	4	50	3.96	−44	−60	0
	Planum Temporale	LH																29	3.66	−54	−42	18
	Superior Temporal Gyrus posterior	LH	121	4.07	−62	−38	2						38	3.28	−66	−34	0					
	Temporal Fusiform Cortex posterior	LH	330	6.76	−36	−50	−28	56	3.65	−44	−44	−18	295	5.69	−36	−50	−28	257	5.75	−36	−48	−28
		RH	81	5.22	36	−42	−30						80	4.97	34	−40	−30	70	4.05	36	−42	−30
	Temporal Occipital Fusiform Cortex	LH	686	6.56	−36	−50	−26	466	5	−42	−58	−26	636	6.3	−42	−60	−16	677	6.23	−42	−62	−22
		RH	671	6.52	34	−50	−26	358	5.34	40	−58	−26	627	5.82	36	−52	−26	655	5.45	40	−60	−22
	Temporal Pole	LH	419	5.02	−54	14	−6						401	5.68	−52	12	−4	277	3.99	−54	12	−10
		RH	176	3.93	58	14	−8						134	3.92	52	16	−4					
Occipital	Cuneal Cortex	RH	43	3.85	2	−88	46															
	Intracalcarine Cortex	LH	164	4.4	−18	−88	2	41	4	−18	−88	2	66	4.35	−18	−88	2	268	4.82	−18	−88	2
		RH	84	2.93	10	−72	4	30	4.38	12	−88	0	57	4.51	12	−88	0	185	4.21	12	−88	0
			61	3.91	12	−88	0															
	Lateral Occipital Cortex inferior	LH	1983	7.21	−36	−86	−8	1437	6.2	−30	−86	−8	1886	6.84	−30	−86	−10	2058	7.21	−30	−84	−8
		RH	1966	7.09	32	−86	−14	1474	6.15	30	−86	−6	1833	6.46	32	−84	−6	1839	6.33	38	−86	−8
	Lateral Occipital Cortex superior	LH	220	4.32	−32	−88	8	617	4.57	−30	−66	36	1867	5.41	−28	−66	40	656	4.93	−28	−60	50
			25	3.98	−4	−84	52	63	3.81	−26	−88	8	170	4.4	−30	−88	10	161	4.4	−22	−90	8
		RH	218	4.93	26	−88	6	129	4.81	26	−88	6	161	5.1	26	−88	6	280	4.94	26	−88	6
								114	3.71	30	−58	50	37	4.1	14	−70	52	127	3.29	30	−58	52
	Occipital Fusiform Gyrus	LH	1421	6.81	−30	−86	−18	1246	6.41	−20	−90	−10	1417	6.69	−28	−86	−10	1504	7.08	−28	−84	−8
		RH	1380	7.48	24	−88	−12	1054	5.96	30	−84	−6	1350	6.47	20	−88	−10	1382	6.3	22	−88	−12
	Occipital Pole	LH	1918	6.92	−16	−94	−8	1783	6.53	−18	−92	−10	1680	7.31	−26	−92	−2	2103	7.06	−18	−94	−4
		RH	1891	6.83	24	−90	−6	1460	6.88	14	−92	−8	1322	6.28	22	−94	0	1834	6.9	24	−92	−4
Medial	Amygdala	LH	156	4.22	−20	−10	−12						71	3.36	−16	−10	−14					
		RH	50	3.45	14	0	−16															
	Hippocampus	LH	239	4.25	−22	−26	−8						78	3.83	−22	−30	−6					
		RH	39	3.78	20	−28	−8															
	Insular Cortex	LH	566	5.77	−34	22	0						368	5.31	−32	22	−2	204	3.45	−32	−6	12
													144	3.8	−38	−4	12					
		RH	326	5.22	32	20	0						288	4.74	40	18	−4					
	Parahippocampal Gyrus anterior	LH	51	3.39	−16	−6	−18															
	Parahippocampal Gyrus posterior	LH	317	4.28	−10	−32	−8						103	3.83	−24	−32	−6					
		RH	132	3.78	20	−28	−8															
Basal Ganglia	Caudate	LH	364	5.74	−14	2	14						202	4.73	−10	−2	12					
		RH	189	5.01	12	12	4															
	Pallidum	LH	173	4.72	−12	4	−2						103	3.84	−16	6	0	34	3.05	−24	−16	0
		RH	88	3.91	16	8	−2															
	Putamen	LH	600	4.59	−16	10	−2						576	4.45	−26	4	−4	116	3.55	−26	−10	12
		RH	338	4.45	18	6	−10						384	3.64	24	10	4					
Midbrain	Thalamus	LH	1039	5.31	−10	−2	10						684	4.84	−10	−4	10	194	3.54	−10	−18	4
		RH	663	5.1	2	−20	10						396	4.39	2	−14	14					

Table 4
Activation coordinates for contrasts between aloud, silent, and sensorimotor control conditions during the test phase, constrained to areas that were active for the same contrasts during the study phase.

Conjunction Study & Test Phase			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	ROI	Hemisphere	Cluster size (mm3)	Max z	x	y	z	Cluster size (mm3)	Max z	x	y	z	Cluster size (mm3)	Max z	x	y	z
Frontal	Middle Frontal Gyrus	LH											25	3.31	−32	0	46
	Precentral Gyrus	LH	81	3.56	−48	4	16						25	2.78	−46	−6	40
Parietal	Central Opercular Cortex	LH	60	3.5	−42	2	6										
	Postcentral Gyrus	LH											25	3.06	−44	−20	46
Temporal	Lingual Gyrus	LH	60	2.79	0	−80	−22										
		RH	137	3.16	4	−80	−20										
			92	3.4	14	−60	−14										
	Superior Temporal Gyrus posterior	LH	41	3.22	−60	−38	2										
Occipital	Lateral Occipital Cortex superior	LH											27	2.88	−30	−64	46
	Occipital Fusiform Gyrus	RH	34	3.28	30	−72	−26										
			25	3.05	14	−84	−24										
Medial	Insular Cortex	LH	40	3.65	−40	2	6										

then compared between the two halves to gain an estimate of internal consistency. The (Spearman-Brown corrected) mean split-half internal consistency for combined "old" judgments was r_{SB} = . 0.35, 95% CI [− 0.06, 0.65]; for recollection r_{SB} = 0.44, 95% CI [0.07, 0.72].[5]

Neuroimaging

Study phase

Activation for each condition was first contrasted with fixation baseline, to identify the broad networks engaged in each condition. Details of these results are provided in Table 1. All three conditions—aloud, silent, and sensorimotor control—were associated with extensive bilateral cortical activation (more extensive in the left hemisphere) in all lobes of the cortex as well as the midbrain, but the silent condition was associated with relatively less extensive activity. Similar regions were activated in all three conditions but an absence of activation in the silent condition was notable in the following regions: auditory processing regions of the superior and middle temporal gyri (including Heschl's gyrus and the planum polare); the hippocampus, parahippocampal gyri, basal ganglia and amygdalae; posterior medial cortical regions including the supra-calcarine cortex, parietal operculum, and posterior cingulate gyri.

The contrasts between conditions, shown in Fig. 2 and Table 2, focused on our research questions. We first consider brain areas activated significantly more in the aloud condition than in the control condition (which was similar in terms of motor activity and auditory perception of self-generated speech). Reading the study words aloud, compared to saying the word "check" while reading them, yielded greater activation along the superior temporal cortex bilaterally, including areas consistent with primary and secondary auditory cortices (including Heschl's gyrus and the planum temporale) as well as more anterior regions often associated with speech processing. Activation was also found in the inferior motor cortex region (pre- and post-central gyri and central opercula), consistent with areas associated with speech articulators; this activation was bilateral but more extensive in the right hemisphere.

The contrasts for the aloud and control conditions against the silent condition yielded relatively more extensive differences, but these two contrasts yielded generally similar patterns of activation as shown in Fig. 2, with details provided in Table 2. For both contrasts, the central sulci (primary motor/sensory cortices) and superior temporal gyri were more strongly and extensively activated bilaterally, relative to the aloud versus control contrast. Activation in these contrasts also extended into additional areas including the frontal lobes (bilateral IFG, SFG, SMA, left MFG), inferior and superior parietal regions (bilateral SMG; left SPL and left AG were activated only for the control-silent contrast), and temporal-occipital areas including the middle and inferior temporal gyri, lateral occipital cortex, fusiform and lingual gyri. Extensive medial and subcortical activation was also obtained, including in the hippocampi, parahippocampal gyri, amygdalae, cingulate gyri (anterior and posterior), and basal ganglia (putamen, pallidum; caudate for the aloud-silent contrast only).

Given that the behavioral production effect was characterized by better memory for aloud items than either silent or control items, we performed a conjunction analysis to identify the brain regions that were significantly activated in the contrasts of the aloud condition against each of the other conditions; this analysis is shown in the bottom right panel of Fig. 2. The brain areas consistently associated with the "production effect" contrasts during study were those identified in the aloud versus control contrast—motor and auditory cortices—confirming that these were a subset of the regions identified in the control versus silent contrast.

[5] We did not compute split-half reliability for familiarity because calculating the familiarity-based production effect requires the exclusion of all trials with a "recollect" response. Therefore, reliability calculations for familiarity would be based on an extremely low number of trials (as low as 6 trials per half for some participants). More generally, our reported estimates may be imprecise, and should therefore be viewed with caution, again owing to the low number of trials involved in these calculations (15 trials for each condition in each half). Indeed, this experiment was not designed to accurately assess reliability and split-half calculations were undertaken only because no other reports of reliability for this measure are available within the literature. The robustness of our behavioural results is further supported by the fact that the experiment was sufficiently powered to detect expected effect sizes, and that our results are entirely consistent with prior literature.

Table 5
Activation coordinates for aloud, silent, and sensorimotor control conditions relative to foil items during the test phase, constrained to areas that were activated by the minuend condition relative to baseline at study.

Conjunction Study & Test Phase				Aloud-Foil				Silent-Foil					Control-Foil				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	229	3.8	−2	22	34						109	3.45	−4	24	32
		RH	42	3.44	2	22	34										
	Frontal Operculum Cortex	LH	231	4.06	−42	16	4						87	3.3	−50	12	−4
	Frontal Orbital Cortex	LH	256	4.05	−38	22	−4						322	3.8	−32	24	−4
	Frontal Pole	LH	83	4.57	−44	40	6	32	3.25	−42	44	8	1459	4.85	−34	60	4
	Inferior Frontal Gyrus pars opercularis	LH	491	4.27	−54	14	22						771	4.09	−34	16	22
	Inferior Frontal Gyrus pars triangularis	LH	252	3.92	−44	36	8	30	2.92	−48	28	22	726	4.07	−46	34	16
	Middle Frontal Gyrus	LH	130	4.56	−38	0	50	431	4.49	−44	22	34	1742	4.65	−48	32	24
		RH											58	2.81	32	4	64
	Paracingulate Gyrus	LH	221	4.28	0	22	40						374	4.22	−4	28	38
		RH	50	3.9	2	20	40										
	Precentral Gyrus	LH	1080	4.36	−40	−2	50						848	4.01	−36	4	32
			81	3.13	−4	−14	76										
	Superior Frontal Gyrus	LH	400	4.51	−2	10	66						118	3.97	−6	24	48
		RH	93	4.07	2	10	62										
	Supplementary Motor Area	LH	374	4.59	−2	8	66										
		RH	126	4.06	2	8	62										
Parietal	Angular Gyrus	LH						171	4.26	−40	−56	40	387	5.2	−38	−56	38
	Central Opercular Cortex	LH	70	3.59	−44	4	14										
	Cingulate Gyrus posterior	LH	67	4.12	0	−40	2						50	3.49	0	−32	6
		RH	42	3.77	2	−40	2										
	Postcentral Gyrus	LH	125	3.5	−4	−50	72						73	3.3	−42	−22	48
			29	3.35	−38	−30	46						60	3.15	−8	−48	68
	Precuneous Cortex	LH	183	3.49	−2	−80	58						483	4.09	−8	−60	60
		RH											26	3.12	6	−64	58
	Superior Parietal Lobule	LH											315	4.61	−34	−58	40
													26	3.19	−12	−58	60
	Supramarginal Gyrus anterior	LH											90	4.24	−50	−40	48
	Supramarginal Gyrus posterior	LH						40	3.53	−46	−48	46	494	4.53	−42	−52	42
Temporal	Inferior Temporal Gyrus posterior	LH											52	3.06	−68	−44	−18
	Inferior Temporal Gyrus temporooccipital	LH	419	4.1	−56	−54	−20						30	3.07	−64	−48	−14
		RH	55	3.38	54	−58	−28										
	Lingual Gyrus	LH	132	3.79	−2	−82	−16										
		RH	337	4.75	6	−78	−24						47	3.96	6	−78	−26
	Middle Temporal Gyrus posterior	LH	127	4.26	−62	−34	−2						444	4.01	−68	−40	−4
	Middle Temporal Gyrus temporooccipital	LH	94	3.7	−60	−46	4						187	3.53	−58	−44	−8
	Superior Temporal Gyrus posterior	LH	59	4.04	−64	−34	0										
	Temporal Fusiform Cortex posterior	LH	26	3.06	−26	−38	−28										
	Temporal Occipital Fusiform Cortex	LH	126	3.44	−48	−58	−24										
		RH	148	4.17	28	−48	−22										
	Temporal Pole	LH	90	3.91	−56	18	−8						146	3.41	−54	20	−14
Occipital	Cuneal Cortex	RH	31	3.2	4	−92	40										
	Lateral Occipital Cortex inferior	LH	153	3.13	−56	−78	−10						25	3.1	−62	−62	−12
		RH	398	4.3	38	−74	−28						103	3.62	40	−72	−28
	Lateral Occipital Cortex superior	LH						505	3.74	−38	−60	42	1350	5.32	−38	−66	46
		RH											30	3.23	14	−72	56
	Occipital Fusiform Gyrus	RH	479	5.01	10	−82	−24						210	3.92	10	−74	−24
	Occipital Pole	RH	60	3.41	2	−92	36										
Medial	Hippocampus	LH	120	3.45	−22	−26	−12										
	Insular Cortex	LH	234	4.39	−36	20	−2						260	4.33	−32	20	−4

(*continued on next page*)

Table 5 (*continued*)

Conjunction Study & Test Phase				Aloud-Foil				Silent-Foil				Control-Foil				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
		RH	38	3	32	20	4						26	2.92	28	16	4
	Parahippocampal Gyrus posterior	LH	231	3.83	−2	−40	0										
		RH	62	3.16	12	−34	−4										
Basal Ganglia	Caudate	LH	330	5.26	−10	4	12						156	4.68	−10	4	8
		RH	166	4.26	10	10	6							3.97	−14	4	2
	Pallidum	LH	64	4.74	−12	4	0						39	3.55	−16	8	−2
	Putamen	LH	111	3.84	−16	8	−2						106	3.19	−32	−14	−8
													59	2.77	24	12	2
		RH											26	4.41	−10	−2	10
Midbrain	Thalamus	LH	720	4.84	−2	−18	12						338	3.2	−16	−20	10
													40	3.73	2	−16	14
		RH	345	4.44	2	−18	12						215				

Table 6
Activation coordinates from the Aloud-baseline contrast during the study phase correlating with behavioural performance (recollection success) at test.

Study Phase Lobe	ROI	Hemi	Aloud-baseline vs "recollect" accuracy				
			Cluster size (mm3)	Max z	x	y	z
Frontal	Precentral Gyrus	LH	41	3.62	−56	2	8
Parietal	Central Opercular Cortex	LH	27	3.44	−54	2	6
	Parietal Operculum Cortex	LH	27	3.56	−48	−30	18

Representational similarity analysis (RSA)

Our RSA investigation of study phase data did not reveal any significant differences between conditions for individual study item activation patterns. However, we did observe some interesting non-significant trends, detailed in Table S2. Aloud items exhibited higher correlations with a phonological model in frontal ROIs (left SMA, right IFG and right precentral gyrus) when compared to silent items, and in temporal ROIs (left posterior ITG and MTG) as well as the occipital pole when compared to control items.

Test phase

Because our primary interest with respect to the test phase concerned activation reflecting encoding processes (i.e., reinstatement), activation from all contrasts at test was masked with the same contrasts at study (with the exception of the foil–baseline contrast). Non-constrained results from the test phase are described in supplementary materials; activation coordinates are reported in Tables S3, S4, and S5.

Activation for each type of item at test relative to fixation is reported in Table 3. Relative to baseline, all item types elicited extensive activation across all lobes of the cerebral cortex and subcortical regions. Notably, much less activation was obtained in the test phase, relative to the study phase, in superior temporal lobe regions associated with auditory processing.

Contrasts between conditions are shown in Fig. 3 and Table 4. No differences were observed for the aloud-control contrast. For the aloud-silent contrast, activation was present in inferior motor cortex (left precentral gyrus and central operculum) and temporo-occipital (bilateral lingual gyrus, left posterior STG) cortices associated with vocalization and auditory processing. This contrast also yielded focal activation in the fusiform gyrus on the inferior temporal lobe. The control-silent yielded a similar, though less extensive pattern of activation, including clusters in left MFG, pre- and postcentral gyri, and superior lateral occipital cortex.

Additional contrasts were made for each study condition relative to the recognition test foils. The contrasts are shown in Fig. 4 and Table 5. Notably, activation elicited by aloud and control conditions relative to foil yielded bilateral activation (though more extensive in the left hemisphere) in frontal (SFG, MFG, IFG), sensorimotor (pre- and postcentral gyri), and temporal cortices (MFG, temporal pole). Moreover, the aloud-foil contrast elicited more extensive activation in frontal, temporal and parietal cortices whereas activation for the control-foil contrast was more extensive in the parietal lobe (IPL, SMG). The silent-foil contrast yielded more focal activation, with significant clusters in frontal (IFG and MFG), parietal (AG, SMG), and occipital (superior lateral occipital cortex) cortices in the left hemisphere. No areas were commonly activated by all three studied conditions relative to foil items (but see supplementary materials for a description of commonly activated areas when test phase results were not constrained to areas activated in the study phase).

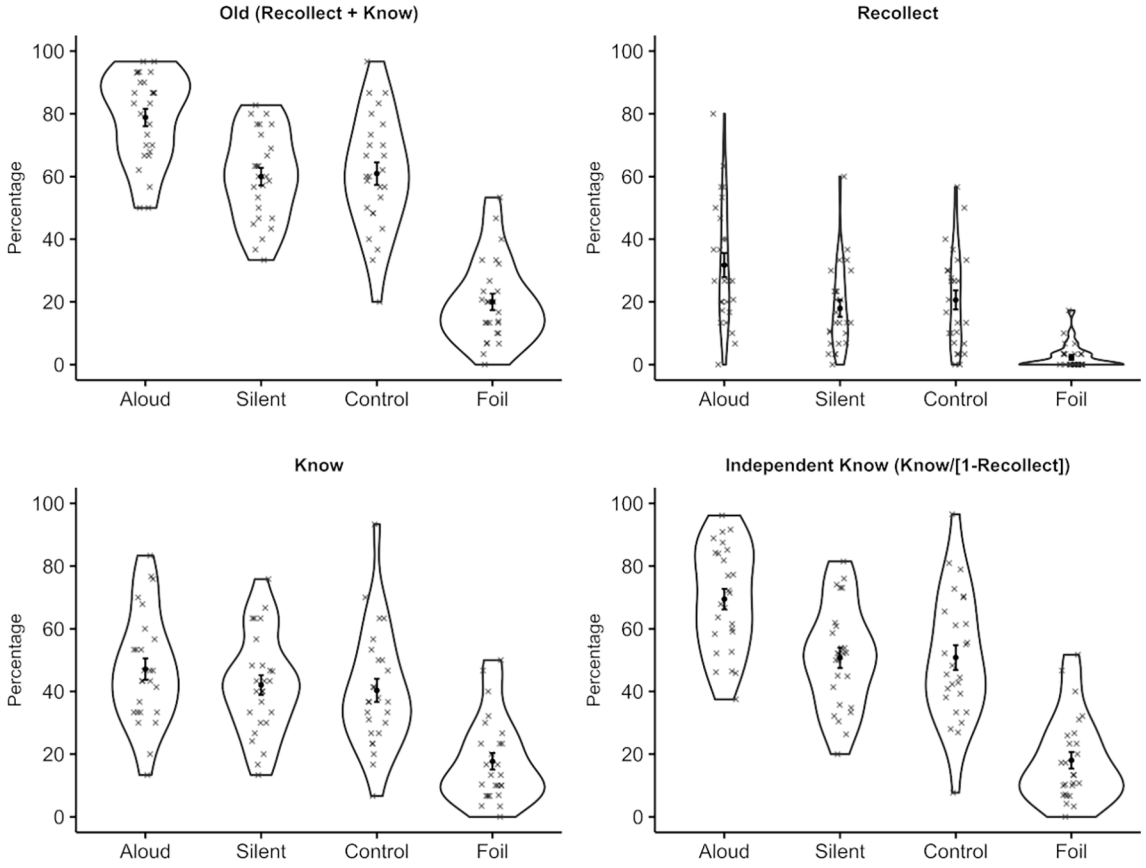


Fig. 1. Mean old responses and separate recollect, know, and independent know responses (%) as a function of item type (aloud, silent, control, foil). Violin plots and X's indicate the distribution of individual participant means. Fitted circles reflect the empirical means; error bars reflect the standard error of the mean. X's have been jittered in the horizontal plane to make them more easily distinguishable.

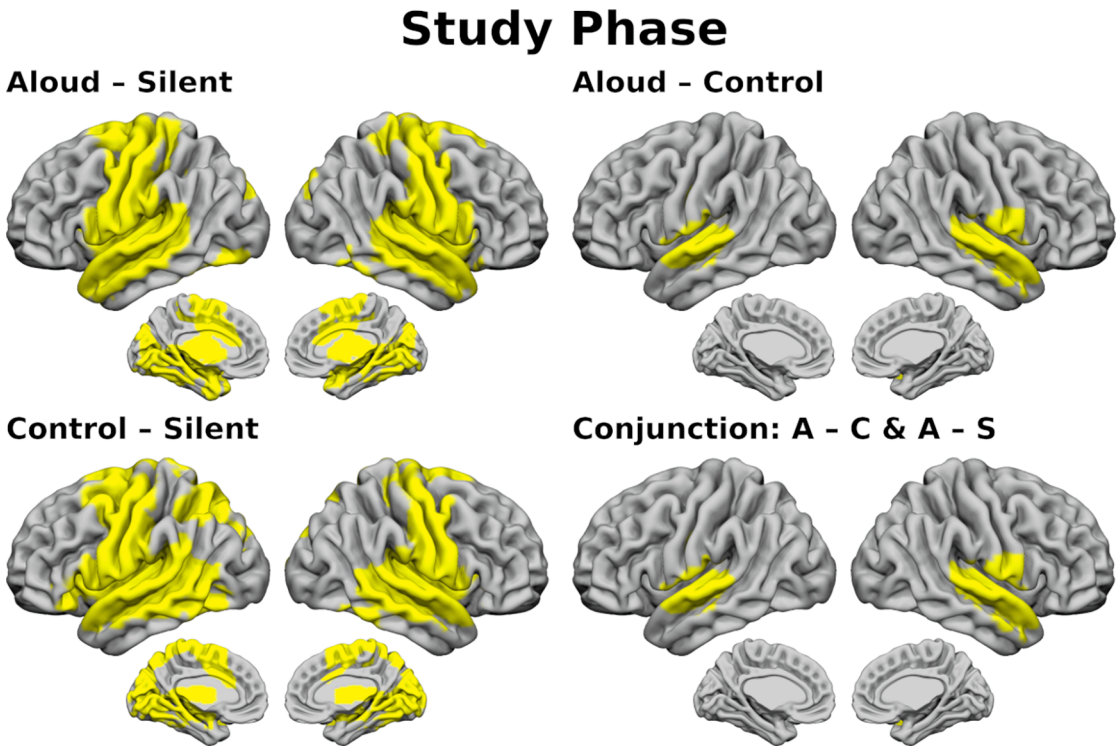


Fig. 2. Differences in activation between conditions in the study phase. Contrasts show aloud items relative to silent items (top left panel), aloud items relative to sensorimotor control items (top right), sensorimotor control items relative to silent items (bottom left), and conjunction of aloud relative to sensorimotor control and aloud relative to silent (bottom right).

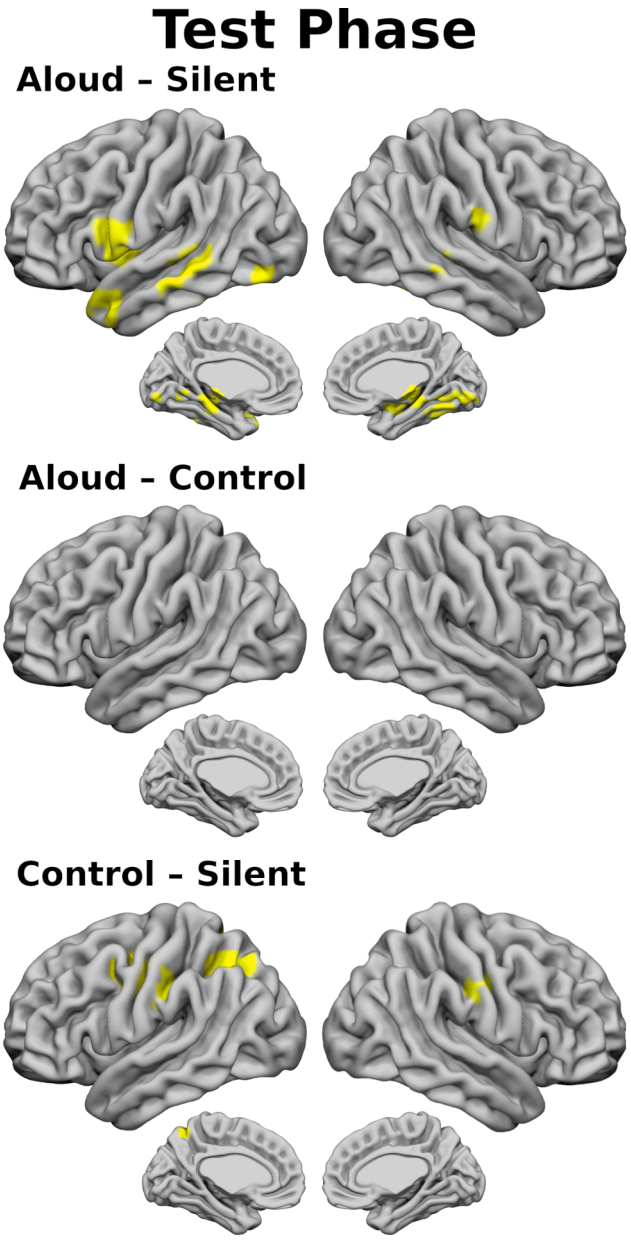


Fig. 3. Differences in activation between conditions in the test phase, constrained to areas showing activation for the same contrasts in the study phase. Contrasts show aloud items relative to silent items (top panel), aloud items relative to sensorimotor control items (middle panel), sensorimotor control items relative to silent items (bottom panel).

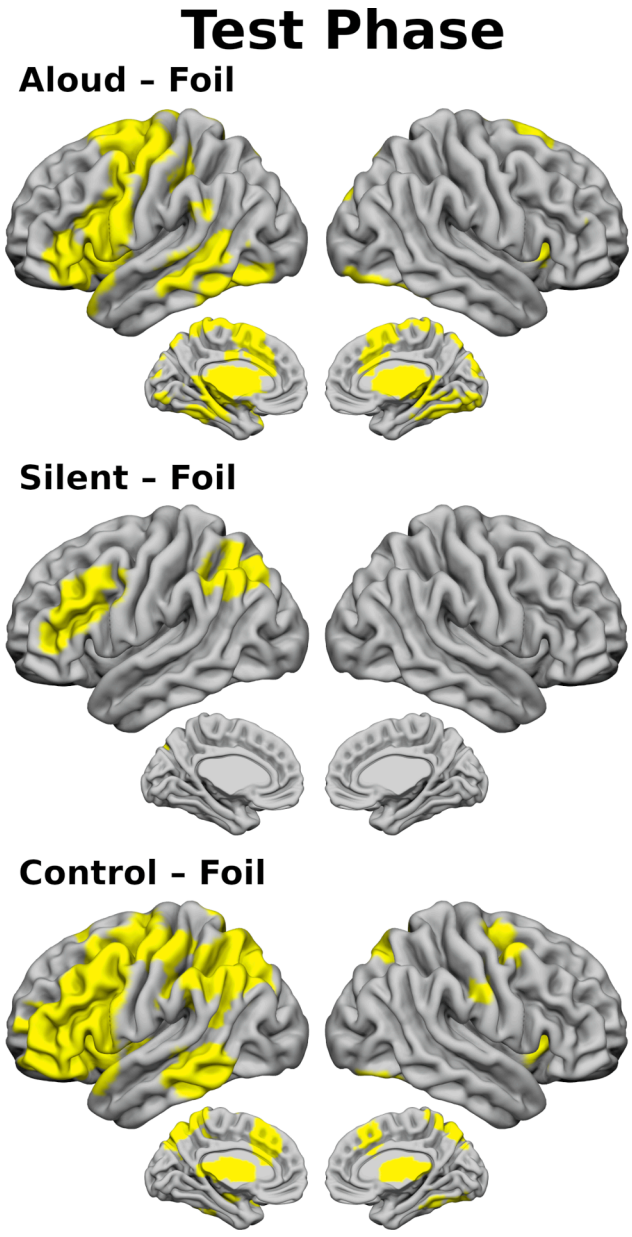


Fig. 4. Differences in activation between foil items and all other conditions in the test phase. Contrasts show aloud items relative to foil (top panel), silent items relative to foil (middle panel), and sensorimotor control items relative to foil (bottom panel).

Brain-Behavior correlations

We also investigated correlations between brain activation and both (1) overall performance in each condition (aloud, control, silent) as indexed by recollect and know judgments (separately), as well as for the combined 'old' (recollect + know) judgments, each corrected for false alarms; and, (2) the behavioural production effect as indexed by recollect and know judgments (separately), as well as for the combined 'old' (recollect + know) judgments. Significant correlations from these analyses (surviving multiple comparison correction) are shown in Fig. 5 and Table 6.

With respect to successful recollection, activation in the aloud-baseline contrast during the study phase significantly correlated with recollect judgments in left inferior motor cortex regions (precentral gyrus, central and parietal opercular cortices). No other significant correlations were present with respect to the aloud or silent conditions relative to baseline. Moreover, there were no significant correlations with respect to either the aloud-silent or aloud-control contrast.

Discussion

The present study is the first to use fMRI to characterize the neural mechanisms giving rise to the production effect. Participants studied subsets of words aloud, silently, or by making a non-unique verbal "check" response (sensorimotor control condition), followed by a recognition memory test. With respect to behavioural findings, a production effect was obtained, with greater recollection and familiarity ratings for the aloud items than for the silent items. These findings replicate earlier work in this area (Fawcett & Ozubko, 2016; Ozubko et al., 2012). Our primary focus, however, was in understanding the neural correlates of the production effect. In this respect, a distinctiveness-based account predicted that reading aloud, relative to

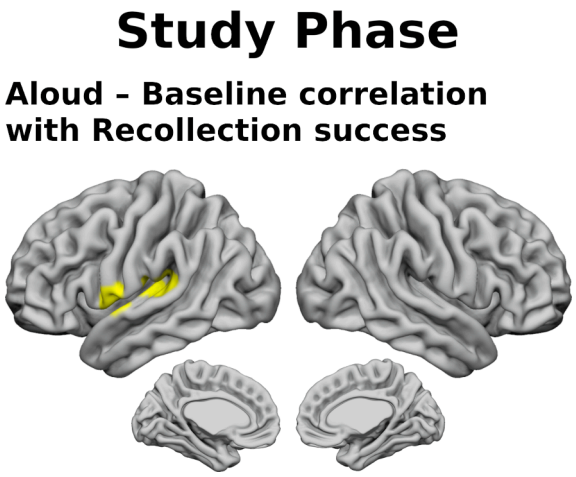


Fig. 5. Coloring indicates activation from the aloud–baseline contrast during the study phase that significantly correlated with recollection success for aloud items at test.

the other conditions, would result in stronger activation of sensorimotor and phonological regions both at study (reflecting the encoding of these features) and test (reflecting their retrieval). This account was supported by the data, as we summarize below according to the experimental phase.

Study phase

Contrasting the three encoding conditions against the fixation baseline revealed activation in brain regions typically associated with encoding in verbal memory tasks, along with regions reflecting the relative sensorimotor demands of the tasks. These activations included inferior frontal (including IFG and MFG), premotor, and parietal (IPS and SMG) regions. Occipital and inferior temporal cortex were also activated (consistent with visual presentation of words), as was the posterior/middle superior temporal sulcus region (consistent with lexical processing). The conditions involving motor speech output (aloud and control) activated motor and somatosensory cortex along the central sulcus, basal ganglia (also associated with sensorimotor control), and auditory processing regions in the superior temporal gyrus and extending into inferior parietal and middle temporal regions. Hippocampal and parahippocampal gyri were also more activated in the aloud and control conditions, which we attribute to greater allocation of attention in those conditions (see our discussion of study contrasts).

Our primary goal was to identify brain regions that showed differential activation for the contrasts between the aloud condition and the two other conditions – those revealing the neural basis of the production effect. The aloud–silent contrast elicited more extensive activation than the aloud–control contrast, but recognition did not differ credibly between the silent and control conditions. Therefore, differences between these contrasts are not likely due to factors driving the production effect. In particular, the stronger activation in motor and auditory cortices for the aloud and control conditions likely reflected these tasks' recruitment of motor speech production and auditory perception by both of these tasks.

For this reason, we focus on areas that were consistent in the contrasts of the aloud condition with the other two conditions. Most extensively, the superior temporal lobe bilaterally was most activated in the aloud condition, from the planum temporale through Heschl's gyrus (primary auditory cortex) to anterior regions commonly associated with speech processing (Hickok & Poeppel, 2016; Venezia et al., 2017), including the superior temporal sulcus and part of the right middle temporal gyrus. Stronger activation was also obtained for aloud items in the inferior parts of the precentral and postcentral gyri (bilaterally although more extensive in the right hemisphere) – areas involved in the motor control of speech. This pattern is consistent with a distinctiveness account in which the articulatory and sensory (auditory and somatosensory) experiences that occur during production are incorporated into the production record. Supporting this interpretation, activation in inferior motor regions correlated with recollection success for the aloud condition (relative to baseline).

Surprisingly, we did not observe correlations with respect to either contrast (aloud–silent or aloud–control) or the behavioural production effect. This may be because activation in the silent and control conditions are poorly correlated with test performance (indeed, neither the silent–baseline nor control–baseline contrasts yielded significant correlations with test performance on trials of their respective condition). Given that activation in the silent condition contributes to the aloud–silent contrast, this may have introduced noise (uncorrelated variance) which masked potential correlations with the behavioural production effect. A similar account may explain why the aloud–control contrast did not correlate with a behavioural production effect defined as performance on aloud trials minus control trials. Alternatively, the absence of correlations with between-conditions contrasts may be due to a lack of sensitivity, given that the magnitude of between-condition differences are inherently lower than for contrasts of conditions involving stimulus presentation and/or motor responses relative to a resting baseline. More generally, these correlation analyses should be viewed with some caution; they were largely exploratory, and it is also possible that the absence of correlations with respect to the production effect and/or either contrast was due to limited power for an fMRI investigation of individual differences (Dubois & Adolphs, 2016). As such, the significant correlation between aloud–baseline activation and recollection accuracy might be viewed as providing greater confidence in a distinctiveness account of our main findings (whereby sensorimotor activation facilitates later retrieval of aloud items), but may not be immediately informative as to the neural mechanisms underlying the production effect. This finding warrants replication with a larger sample size.

Finally, our multivariate analysis (RSA) indicated some interesting non-significant trends whereby activation for aloud items was more distinctive—evidenced by higher correlations with a phonological model—in areas associated with articulation (SMA, IFG, precentral gyrus) when compared to silent items; and areas associated with lexical processing (ITG, MTG) when compared to control items. Although these between-condition differences were non-significant, it is important to recognize that the study was not designed with this analysis in mind, and that these were exploratory post hoc analyses. As such, they provide further — however tentative — support for a distinctiveness account and warrant further investigation by future studies, perhaps with designs specifically tailored to MVPA (e.g., Zeithamova, de Araujo Sanchez, & Adke, 2017).

Our results are also consistent with the possibility that the production effect arises in part from increased attentional engagement or supplementary processing during aloud trials. Task-relevant activation in sensorimotor areas was greater for the aloud relative to the control condition, congruent with attentional up-regulation on aloud trials (e.g., Johansen-Berg & Matthews, 2002; Rinne et al., 2005; Rowe, Friston, Frackowiak, & Passingham, 2002). However, we cannot rule out the possibility that sensorimotor activation in the control condition was muted due to its repetitive nature. Moreover, activation was present in IFG and superior temporal gyrus in both the aloud and control conditions relative to silent. With respect to semantic processing, a meta-analysis of brain networks related to semantic comprehension of spoken and written language implicated both the IFG and superior temporal gyrus (Rodd, Vitello, Woollams, & Adank, 2015). Therefore, enhanced encoding and semantic processing of aloud items may generate more stable memory representations, facilitating later recollection.

Medial temporal regions such as the hippocampus were more active

during aloud than silent trials but, interestingly, they were also more active during control trials—despite recognition being similar in the control and silent conditions. Moreover, activation in medial temporal regions at study did not correlate with later recognition. Given that the hippocampus is often associated with successful encoding, and is also known to be modulated by attentional manipulations, this lack of correlation was unexpected. This suggests that encoding was enhanced for both aloud and control trials, but proved to be of little benefit for the control trials because the dominant feature in that episode (i.e., having said "check") was not deemed to be as diagnostic of prior study as retrieval of having said the actual test item aloud. Indeed, hippocampal activation in the aloud condition might reflect greater attention to and encoding of the stimuli, whereas hippocampal activation in the control condition might reflect greater attention to and encoding of the response. This possibility warrants further exploration.

Test phase

Surprisingly, we did not find differences in brain activation between the aloud and control conditions at test. However, activation of areas in the aloud–silent and control–silent contrasts that were, critically, also active during encoding may indicate reinstatement of task-related processes. In particular, activation of areas associated with articulation and auditory processing (somatosensory cortex and posterior STG) may reflect recollection of speech production for both aloud and control items. Importantly, such retrieval would only be diagnostic at test for aloud items; recollection of speaking a nonspecific word ("check") was likely insufficient to differentiate specific words from one another (evidenced by the absence of a behavioural production effect for the control condition). Moreover, activation of the fusiform gyrus (which houses the visual word form area) in the aloud–silent contrast may reflect more vivid recollection of reading the word during the study phase.

Conclusion

Producing items during study, particularly by reading them aloud, provides a simple and effective means of enhancing memory (MacLeod & Bodner, 2017). Our fMRI study explored the neural basis of the production effect. Our results are compatible with the dominant distinctiveness account, in demonstrating greater activation of primary sensorimotor cortex (associated with articulation) and auditory cortex (associated with perception) for produced than non-produced words during encoding. This account is further supported by our findings that activation in these regions correlated with later recognition only for produced items, and was somewhat more distinctive for aloud compared to silent and control items. However, our data also suggest that participants may be more engaged during aloud than silent trials. For example, they showed heightened activation of task-relevant regions on aloud trials, and greater recruitment of areas implicated in semantic processing. These differences emerged in analyses based only on items that were later correctly recognized, thus they were not artefacts of the aloud condition yielding proportionately better memory performance. Future studies should investigate these patterns of activation in more detail, for example by using network-based connectivity and/or designs more tailored to MVPA.

CRediT authorship contribution statement

Lyam M. Bailey: Conceptualization, Formal analysis, Data curation, Writing - review & editing, Visualization. **Glen E. Bodner:** Conceptualization, Methodology, Writing - review & editing. **Heath E. Matheson:** Conceptualization, Formal analysis. **Brandie M. Stewart:** Investigation, Writing - review & editing. **Kyle Roddick:** Conceptualization, Methodology, Investigation. **Kiera O'Neil:** Investigation. **Maria Simmons:** Investigation. **Angela M. Lambert:** Investigation. **Olave E. Krigolson:** Conceptualization, Methodology. **Aaron J. Newman:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Jonathan M. Fawcett:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

We are grateful to the following people for their assistance in this study: Lauren Dunphy, Morgan Johnson, Greg MacLean, Matthew Rodgers, and Sarah Sullivan. Funding was provided by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to GB, OEK, JF and AN, and from the Nova Scotia Health Research Foundation to OEK and AJN.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2021.105757.

References

Andersson, J. L. R., Jenkinson M., & Smith, S. M. (2007a) Non-linear optimisation. FMRIB technical report TR07JA1.
Andersson J. L. R., Jenkinson M., & Smith S. M. (2007b). Non-linear registration, aka Spatial normalisation. FMRIB technical report TR07JA2.
Astafiev, S. V., Shulman, G. L., & Corbetta, M. (2006). Visuospatial reorienting signals in the human temporo-parietal junction are independent of response selection. *European Journal of Neuroscience, 23*(2), 591–596. https://doi.org/10.1111/j.1460-9568.2005.04573.x.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Special Issue: Emerging Data Analysis, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.
Bodner, G. E., Taikh, A., & Fawcett, J. M. (2014). Assessing the costs and benefits of production in recognition. *Psychonomic Bulletin & Review, 21*(1), 149–154.
Bürkner, P.-C. (2017a). Advanced Bayesian multilevel modeling with the R package brms. ArXiv Preprint ArXiv:1705.11123.
Bürkner, P.-C. (2017b). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 1*(1), 2017. Retrieved from https://www.jstatsoft.org/v080/i01.
Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0.
Ciaramelli, E., Grady, C., Levine, B., Ween, J., & Moscovitch, M. (2010). Top-Down and Bottom-Up Attention to Memory Are Dissociated in Posterior Parietal Cortex: Neuroimaging and Neuropsychological Evidence. *The Journal of Neuroscience, 30*(14), 4943. https://doi.org/10.1523/JNEUROSCI.1209-09.2010.
Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26*(3), 341–361. https://doi.org/10.1016/0749-596X(87)90118-5.
Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201.
Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8*(2–3), 109–114.
Dale, A. M., Greve, D. N., & Burock, M. A. (1999, June). Optimal Stimulus sequences for Event-Related fMRI. Presented at the 5th International Conference on Functional Mapping of the Human Brain. Duesseldorf, Germany.
Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological bulletin, 136*(1), 87.
Danker, J. F., Tompary, A., & Davachi, L. (2017). Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cerebral Cortex, 27*(7), 3515–3524.
Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it": Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*(1), 155–161.
Dubois, J., & Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends in cognitive sciences, 20*(6), 425–443.
Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica, 142*(1), 1–5. https://doi.org/10.1016/j.actpsy.2012.10.001.
Fawcett, J. M., Baldwin, M. M., Drakes, D. H. & Willoughby, H. V. (submitted for publication). Production improves recognition and reduces intrusions in between-subject designs: An empirical and meta-analytic investigation.

Fawcett, J. M., Lawrence, M. A., & Taylor, T. L. (2016). The representational consequences of intentional forgetting: Impairments to both the probability and fidelity of long-term memory. *Journal of Experimental Psychology. General, 145*(1), 56–81. https://doi.org/10.1037/xge0000128.

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale, 70*(2), 99–115. https://doi.org/10.1037/cep0000089.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory (Hove, England), 20*(7), 655–666. https://doi.org/10.1080/09658211.2012.693510.

Forrin, N. D., & MacLeod, C. M. (2016). Order information is used to guide recall of long lists: Further evidence for the item-order account. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 70*(2), 125.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40*(7), 1046–1055. https://doi.org/10.3758/s13421-012-0210-8.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gottfried, J. A., Smith, A. P., Rugg, M. D., & Dolan, R. J. (2004). Remembrance of odors past: Human olfactory cortex in cross-modal recognition memory. *Neuron, 42*(4), 687–695.

Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale, 70*(2), 139–146. https://doi.org/10.1037/cep0000093.

Hickok, G., & Poeppel, D. (2016). Chapter 25—Neural Basis of Speech Perception. In G. Hickok & S. L. Small (Eds.), Neurobiology of Language (pp. 299–310). https://doi.org/10.1016/B978-0-12-407794-2.00025-0.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. *Distinctiveness and Memory*, 3–25.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. *Scientific Approaches to Consciousness*, 13–47.

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 70*(2), 154.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage, 17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132.

Jenkinson, M., & Smith, S. M. (2001). A Global Optimisation Method for Robust Affine Registration of Brain Images. *Medical Image Analysis, 5:2*(143–156).

Johansen-Berg, H., & Matthews, P. (2002). Attention to movement modulates activity in sensori-motor areas, including primary motor cortex. *Experimental Brain Research, 142*(1), 13–24. https://doi.org/10.1007/s00221-001-0905-8.

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. *R package version*, (7).

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage, 54*(3), 2446–2461. https://doi.org/10.1016/j.neuroimage.2010.09.045.

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. (Version 1.3.0). Retrieved from https://CRAN.R-project.org/package=emmeans.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*(4).

Lin, O. Y., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66*(3), 212.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica, 98*(2), 291–310. https://doi.org/10.1016/S0001-6918(97)00047-4.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*(4), 390–395.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 671–685. https://doi.org/10.1037/a0018785.

Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11*(1), 77–95. https://doi.org/10.1016/S0926-6410(00)00066-5.

Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping, 36*(5), 1963–1981. https://doi.org/10.1002/hbm.22749.

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. https://doi.org/10.1016/j.neuroimage.2011.08.076.

Murphy, K. A., Jogia, J., & Talcott, J. B. (2019). On the neural basis of word reading: A meta-analysis of fMRI evidence using activation likelihood estimation. *Journal of Neurolinguistics, 49*, 71–83.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A., & Jonides, J. (2012). A Meta-analysis of Executive Components of Working Memory. *Cerebral Cortex, 23*(2), 264–282. https://doi.org/10.1093/cercor/bhs007.

Newman, A. J., Supalla, T., Fernandez, N., Newport, E. L., & Bavelier, D. (2015). Neural systems supporting linguistic structure, linguistic experience, and symbolic communication in sign language and gesture. *Proceedings of the National Academy of Sciences, 112*(37), 11684. https://doi.org/10.1073/pnas.1510527112.

Newman, A. J., Supalla, T., Hauser, P. C., Newport, E. L., & Bavelier, D. (2010a). Prosodic and narrative processing in American Sign Language: An fMRI study. *NeuroImage, 52*(2), 669–676. https://doi.org/10.1016/j.neuroimage.2010.03.055.

Newman, A. J., Supalla, T., Hauser, P., Newport, E. L., & Bavelier, D. (2010b). Dissociating neural subsystems for grammar by contrasting word order and inflection. *Proceedings of the National Academy of Sciences of the United States of America, 107*(16), 7539–7544. https://doi.org/10.1073/pnas.1003174107.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General, 129*(2), 242.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics, 10*, 27.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functionalneuroimaging studies. *Human Brain Mapping, 25*(1), 46–59. https://doi.org/10.1002/hbm.20131.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1.

Parsons, S. (2020). splithalf: Robust estimates of split half reliability, version 4.0. figshare. *Software*. https://doi.org/10.6084/m9.figshare.11956746.v4.

Peirce, J. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10. https://doi.org/10.3389/neuro.11.010.2008.

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory, 21*(8), 904–915.

R Core Team. (2016). R: A language and environment for statistical computing. (Version 3.5.1). Retrieved from https://www.R-project.org/.

Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jääskeläinen, I. P., Sams, M., & Alho, K. (2005). Modulation of auditory cortex activation by sound presentation rate and attention. *Human Brain Mapping, 26*(2), 94–99. https://doi.org/10.1002/hbm.20123.

Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral cortex, 23*(12), 2818–2828.

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language, 141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., … Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage, 60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050.

Rowe, J., Friston, K., Frackowiak, R., & Passingham, R. (2002). Attention to Action: Specific Modulation of Corticocortical Interactions in Humans. *NeuroImage, 17*(2), 988–998. https://doi.org/10.1006/nimg.2002.1156.

Schultz, H., Tibon, R., LaRocque, K. F., Gagnon, S. A., Wagner, A. D., & Staresina, B. P. (2019). Content tuning in the medial temporal lobe cortex: Voxels that perceive, retrieve. *ENeuro, 6*(5).

Sestieri, C., Shulman, G. L., & Corbetta, M. (2017). The contribution of the human posterior parietal cortex to episodic memory. *Nature Reviews Neuroscience, 18*, 183.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143–155.

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic Reinstatement in the Medial Temporal Lobe. *The Journal of Neuroscience, 32*(50), 18150. https://doi.org/10.1523/JNEUROSCI.4156-12.2012.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061.

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *Neuroimage, 221*, Article 117164.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin, 139*(4), 766–791. https://doi.org/10.1037/a0030266.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Oxford, England: Bureau of Publications, Teachers Co.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne, 26*(1), 1–12. https://doi.org/10.1037/h0080017.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia, 40*(12), 2136–2143.

Varao Sousa, T. L., Carriere, J. S. A., & Smilek, D. (2013). The way we encounter reading material influences how frequently we mind wander. *Frontiers in Psychology, 4*, 892. https://doi.org/10.3389/fpsyg.2013.00892.

Venezia, J. H., Vaden, K. I., Rong, F., Maddox, D., Saberi, K., & Hickok, G. (2017). Auditory, Visual and Audiovisual Speech Processing Streams in Superior Temporal Sulcus. *Frontiers in Human Neuroscience, 11*, 174. https://doi.org/10.3389/fnhum.2017.00174.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., … Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology,

semantics, and sentence processing. *NeuroImage, 30*(4), 1414–1432. https://doi.org/10.1016/j.neuroimage.2005.11.002.

Vigneau, M., Beaucousin, V., Herve, P.-Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage, 54*(1), 577–593. https://doi.org/10.1016/j.neuroimage.2010.07.036.

Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Part Special Issue: What Is the Parietal Lobe Contribution to Human Memory?, 46*(7), 1787–1799. https://doi.org/10.1016/j.neuropsychologia.2008.01.004.

Wagner, S., Sebastian, A., Lieb, K., Tuscher, O., & Tadic, A. (2014). A coordinate-based ALE functional MRI meta-analysis of brain activation during verbal fluency tasks in healthy control subjects. *BMC Neuroscience, 15*, 19. https://doi.org/10.1186/1471-2202-15-19.

Wammes, J. D., Meade, M. E., & Fernandes, M. A. (2016). The drawing effect: Evidence for reliable and robust memory benefits in free recall. *The Quarterly Journal of Experimental Psychology, 69*(9), 1752–1776. https://doi.org/10.1080/17470218.2015.1094494.

Weber, K., Lau, E. F., Stillerman, B., & Kuperberg, G. R. (2016). The yin and the yang of prediction: An fMRI study of semantic predictive processing. *PLoS ONE, 11*(3), Article e0148637.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences, 97*(20), 11125–11129.

Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of cognitive neuroscience, 27*(4), 679–691.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage, 92*, 381–397. https://doi.org/10.1016/j.neuroimage.2014.01.060.

Woolrich, M. W., Ripley, B. D, Brady, J. M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *Neuroimage, 14*(6), 1370–1386.

Yonelinas, Andrew P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864.

Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language, 34*(5), 622–643. https://doi.org/10.1006/jmla.1995.1028.

Zeithamova, D., de Araujo Sanchez, M. A., & Adke, A. (2017). Trial timing and pattern-information analyses of fMRI data. *Neuroimage, 153*, 221–231.

2.12. Supplementary Materials

2.12.1. Supplementary Methods

Single trial estimation. In order to address our multivariate hypotheses, we first extracted item-specific estimates of neural activation using FEAT. Here we used an iterative modelling procedure proposed by Mumford et al. (2012). In brief: for each trial we modelled a GLM with two regressors: one regressor for that trial, and another regressor for all other trials. This entailed high-pass temporal filtering and motion correction, and we rejected runs with excessive motion as described for our univariate analysis. However, due to concerns that the 6 mm spatial smoothing kernel (used in our univariate analysis) might obscure item-specific activation patterns, performed this analysis on data that were not spatially smoothed. Having performed the above procedure, each subject's single-trial estimates were transformed to standard MNI space using the same procedure described for our univariate analysis.

Searchlight analysis (RSA). For each participant, we conducted a whole brain searchlight analysis (spherical searchlights with radius of 3 mm) to compare their neural response patterns with a formal phonological model using the CoSMo library. Within each searchlight, we constructed a 30 x 30 neural dissimilarity matrix (DSM) comprising pairwise correlation distances (1 – Pearson correlation) between the activity patterns for each study item. This was done separately for each condition, such that each DSM contained pairwise distances only for aloud, silent, or control items. Following this, neural DSMs were correlated with a phonological model comprising pairwise edit distances, reflecting phonological dissimilarity, between study items (Figure S1). Phonological edit distances are an analogue to Levenshtein distance, based on phonological features rather than characters (described here: https://readthedocs.org/projects/corpustools/downloads/pdf/v1.1.0/). Phonological edit distances

were obtained from the Irvine Phonotactic Online Dictionary (IPHOD; Vaden, Halpin & Hickok, 2009) using functions from Phonological CorpusTools toolbox (Hall et al., 2017) and additional custom code implemented in Python 3 (Python Software Foundation). Given that items were randomly allocated to conditions for every participant, we computed separate models for every condition/participant. Relative to the rest of our stimulus set, one study item—"uniform"—elicited unusually high distances from all other items and therefore represented an extreme phonological outlier (see Figure S1 for a comparison of a model with and without this item). Therefore, this item (and its associated neural pattern vectors) was excluded from our searchlight analysis to prevent results being skewed by extreme values.

Statistical analysis of searchlight results. To avoid an extremely high number of multiple comparisons associated with analyzing whole brain maps, and to prevent our univariate results from biases our multivariate results, we constrained the output from our searchlight to 11 ROIs relevant to silent and aloud single word reading, identified by Murphy, Jogia and Talcott (2019; see S Table 1). More specifically, for each ROI we generated a spherical mask (radius = 3 voxels) centered on the MNI coordinates for that ROI, and then combined all 11 spheres into a single mask. Searchlight results from each participant were constrained to the combined mask prior to statistical analysis.

Correlation values from the searchlight results were used as input for random effect cluster statistics, implemented with the Cosmo Monte Carlo Cluster Stat function with "maxsum" multiple comparison correction within the CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016). This entailed two Monte Carlo-based permutation t-tests (10,000 iterations): one comparing correlation values from aloud items to those from silent; the other comparing aloud to control. In each test, correlation values (derived from the correlation between the neural DSM and

phonological model) at every voxel were compared between conditions, yielding a *t* value at each voxel. These *t* values were converted to *z* values; each test therefore generated a statistical map with a *z* value at every voxel within our ROIs. Within each map, *z* values > 1.65 are indicative of a statistically significant difference between conditions at the $p < .05$ level for a one-tailed test.

2.12.2. Supplementary Results

Test phase. Activation for each item type at test relative to fixation baseline is reported in Table S3. Relative to baseline all item types elicited extensive activation across all lobes of the cerebral cortex and subcortical regions. Notably, much less activation was obtained in the test phase, relative to the study phase, in superior temporal lobe regions associated with auditory processing. At the same time, activation in the middle and superior frontal gyri, and superior parietal lobule, was more extensive during the test phase than during the study phase across all conditions.

Contrasting activation between conditions yielded more focal activation differences, as shown in Table S4. In contrast to the study phase, no differences were obtained for the aloud-control contrast at test. For the aloud-silent contrast, activation was present in the left IFG pars opercularis and precentral gyrus. As well, significant differences were obtained in IPL bilaterally (but more extensively in the left hemisphere), including the AG, SMG, and LOC. These differences extended in the left hemisphere into the posterior STG and, in both hemispheres, differences were also found in the posterior portions of the inferior and middle temporal gyri. Medially, greater activation for aloud than silent was found in the precuneus and posterior cingulate cortex, as well as the posterior parahippocampal gyrus. Differences for the control-silent contrast were restricted to the IPL bilaterally, and medial regions. The IPL differences again

included the AG and LOC, and were both more restricted in their spatial extent, and more bilateral, than in the aloud-silent contrast. Differences were obtained medially again in the precuneus and posterior cingulate, but not in the parahippocampal gyri.

Additional contrasts were made for each study condition relative to the recognition test foils. We reasoned that activation common to studied items relative to foil items would identify the neural network associated with recognition of familiar words. In turn, this identification allowed us to determine whether activations for aloud items relative to silent and control items represented increased activation within this same network, or areas specific to production. These contrasts are shown in Table S5. The areas commonly activated by studied items that were correctly recognized as such, relative to foil items, included a cluster including the left MFG, IFG, and frontal pole, and a left IPL cluster including the AG, SMG, and LOC, and extending into the SPL and precuneus. Items associated in the study phase with speech production (aloud and control) also showed stronger activation than foils in smaller areas of the homologous right frontal and IPL regions.

Additional contrasts – results and discussion. Given that the current investigation was concerned with the production effect, which is defined as a behavioural advantage for aloud words versus silent words, we mainly focused on brain activity that was preferentially elicited by the aloud reading and sensorimotor control conditions relative to silent reading (i.e., aloud-silent and control-silent fMRI contrasts). However, contrasts in the opposite direction (silent-aloud and silent-control) are arguably interesting as well, because they may reveal brain areas that are preferentially engaged by silent reading. Stated differently, such contrasts may help us better understand neural processes that are increased during silent reading.

The fMRI analyses described on p. 5 of the Methods section also included between-condition contrasts of the silent condition relative to aloud and control reading, during both the study and test phases. Results from these contrasts are not reported in the manuscript proper, given our focus on neural activity relevant to the production effect. Instead, they are reported below.

The silent-aloud contrast in the study phase revealed activation in the left middle frontal gyrus (MFG), left angular gyrus, and the superior portion of lateral occipital cortex (LOC) bilaterally. These results are shown in Table S6. No other contrast (of the four described above) revealed any significant results.

Left MFG has long been associated with speech planning (see Hertrich et al., 2021 for review) while recent work has indicated that this area actively represents goal-relevant information in contexts where response output requires effortful cognitive control (e.g., in a Stroop task where participants must ignore the meaning of a presented word and instead report the colour in which the word is printed; Freund et al., 2021). We suggest that, in our study, left MFG activation may be related to inhibition of (inappropriate) vocal responses on silent trials. Considering that silent trials were intermixed with aloud and sensorimotor control trials during the study phase, such inhibition likely would have been necessary as participants switched between these different conditions from trial to trial.

The angular gyrus receives input from cortices governing sensory perception in multiple modalities and is thought to serve as a hub for multisensory integration during perception, as well being involved in high-level, goal-directed cognition—for example, manipulating or evaluating incoming information, or orienting attention towards goal-relevant features (see Seghier, 2013 for review). More recently, multivariate fMRI work has indicated that this area may also represent

perceptual features of visually presented stimuli, such as colour and/or object category, particularly when those features are relevant to the one's current goal (e.g., Favila et al., 2020). Why this area should be preferentially activated by silent compared to aloud reading is not entirely clear, though it may reflect differential processing of information between the two tasks. For example, silent reading may entail relatively more emphasis (in terms of cognitive processing / attentional allocation) on visual/orthographic features of presented words, resulting in those features being "up-regulated" in the angular gyrus. The role of superior LOC in silent reading is not clear—this area has previously been associated with visual object recognition (e.g., Grill-Spector et al., 2001) but, to our knowledge, not with word reading. Further work may be needed to clarify the functional significance of this finding.

While results for the silent-aloud contrast were much less extensive than those of the opposite contrast (reported in the main results), the fact that this contrast revealed *any* clusters supports the notion that silent reading may entail cognitive processes that are not present (or, at least, present to a lesser degree) during aloud reading. This provides a rather different perspective from what is typically presented in literature surrounding the production effect, which (to our knowledge) only considers processes that are increased / facilitated during aloud compared to silent reading.

2.12.3. Supplementary Tables & Figures

Table S1. MNI152 coordinates for spherical ROIs (taken from Murphy, Jogia & Talcott, 2019) used in RSA, with anatomical labels derived from Oxford-Harvard cortical and subcortical atlases.

MNI coordinates			
x	**y**	**z**	**Anatomical label**
−50	−48	−8	L Inferior Frontal Gyrus, temperoocipital part *
−50	−8	44	L Precentral Gyrus
−40	28	24	L Middle Frontal Gyrus
−54	−16	8	L Superior Temporal Gyrus
−42	−54	−18	L Inferior Frontal Gyrus, temperoocipital part *
−22	−68	48	L Lateral Superior Occipital cortex *
−44	6	26	L Inferior Frontal Gyrus
48	12	24	R Inferior Frontal Gyrus
−24	−98	−4	L Occipital Pole *
−4	−2	56	L Supplementary Motor Area *
34	−56	50	R Superior Parietal Lobule

L = left hemisphere, R = right hemisphere.

* Some labels reported here differ from those reported by Murphy, Jogia & Talcott (2019) due to disagreements between the AAL atlas (reported by those authors) and the Harvard-Oxford atlases, reported here. L inferior frontal gyrus, temperoociptal part was originally reported as L fusiform gyrus; L lateral superior occipital cortex was originally L Precuneus; L occipital pole was originally L inferior occipital gyrus; L supplementary motor area was originally L medial frontal gyrus.

Table S2. Coordinates for clusters of voxels exhibiting higher correlations between neural data and a phonological model, derived from our searchlight analysis, for aloud items relative to silent and control items.

			Aloud - Silent					Aloud - Control				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Inferior Frontal Gyrus pars opercularis	RH	10	0.221	44	10	26					
	Precentral Gyrus	RH	9	0.221	46	8	26					
	Supplementary Motor Area	LH	29	0.364	-8	-4	52					
Temporal	Inferior Temporal Gyrus posterior	LH						2	0.825	-50	-44	-10
	Inferior Temporal Gyrus temporooccipital	LH						2	0.825	-54	-48	-10
	Middle Temporal Gyrus posterior	LH						1	0.825	-50	-42	-8
	Middle Temporal Gyrus temporooccipital	LH						42	0.825	-54	-46	-10
Occipital	Occipital Pole	LH						28	0.462	-22	-102	-2

Table S3. Activation coordinates for each condition relative to baseline during the test phase.

			Aloud-Baseline					Silent-Baseline					Control-Baseline					Foil-Baseline				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	674	6.07	-4	16	38	82	4.6	-2	16	38	805	5.46	-4	16	38	130	4.23	-4	12	40
		RH	219	5.18	4	20	36	27	3.63	2	20	36	257	4.62	2	20	36	103	3.45	6	8	44
			58	3.22	4	4	22															
	Frontal Operculum Cortex	LH	376	5.16	-46	16	0	133	4.21	-48	14	-2	315	5.55	-50	12	-4					
		RH	172	4.24	38	24	0						159	4.16	42	20	-2					
	Frontal Orbital Cortex	LH	1666	5.68	-32	26	-2	271	4.85	-30	26	-6	1265	5.12	-32	24	-4	54	3.08	-46	16	-8
													42	3.54	-18	4	-14					
		RH	632	5.17	32	26	-6						374	4.33	36	24	-4					
													28	2.94	18	4	-14					
	Frontal Pole	LH	5416	5.11	-34	58	4	1664	4.38	-30	54	0	4089	5.01	-42	48	8					
			107	3.07	-10	40	40															

	RH	1623	4.98	52	34	20	288	3.82	46	36	32	579	4.09	24	52	-16					
												573	4.16	44	34	24					
Inferior Frontal Gyrus pars opercularis	LH	1202	5.72	-60	12	20	739	4.61	-50	14	-2	1154	5.83	-50	14	-4	366	3.86	-50	10	28
	RH	550	4.17	50	24	22	29	3.24	50	24	22	78	3.93	50	16	0					
												29	2.76	60	14	24					
												27	2.92	52	12	30					
Inferior Frontal Gyrus pars triangularis	LH	1000	5.55	-52	26	20	414	4.66	-50	28	22	991	4.73	-50	26	24	126	3.33	-46	28	18
	RH	181	5.31	50	30	20	49	3.57	52	28	22	91	4.6	50	30	20					
		36	2.99	44	28	2						51	3.94	52	18	-4					
		34	3.48	50	18	-6															
Middle Frontal Gyrus	LH	3067	5.87	-34	-4	58	1995	5.03	-36	-4	58	2457	5.38	-34	-4	58	431	4.77	-32	-4	56
																	76	3.09	-46	24	24
	RH	1546	5.72	50	30	22	628	4.84	46	26	28	974	5.04	46	30	22	60	3.19	36	-2	60
Paracingulate Gyrus	LH	991	6.69	-2	14	52	503	5.69	-2	18	40	786	6.57	-4	20	42	379	4.91	-2	14	48
	RH	561	5.91	2	12	52	244	4.54	2	18	40	412	5.29	2	20	38	316	4.52	2	10	48

	Precentral Gyrus	LH	3522	6.41	-50	2	34	2258	5.52	-36	-8	58	3462	5.41	-46	4	42	2862	5.24	-52	-2	36
			252	3.96	-6	-14	70														
		RH	1719	4.94	36	-12	62						1539	4.84	32	-8	62	305	3.56	40	-8	60
	Subcallosal Cortex	LH	120	4.94	-6	16	2	25	4	-8	16	-4	48	3.37	-6	16	2					
		RH	81	4.59	14	14	-8															
	Superior Frontal Gyrus	LH	2565	6.39	-2	14	54	865	4.45	-4	10	68	1145	5.45	-4	24	48	165	3.64	-28	-4	62
																		74	4.24	-2	16	52
																		48	3.15	-8	-6	68
		RH	361	5.65	2	14	54	130	3.98	2	12	66	178	4.42	26	-4	48	162	3.38	10	-4	70
			134	4.11	28	-4	54	32	3.56	2	18	52	108	4.33	2	18	52					
	Supplementary Motor Area	LH	826	5.84	-2	2	62	544	4.37	0	2	66	744	4.72	-2	-2	64	503	4.36	0	6	50
		RH	536	5.45	2	6	58	207	3.98	2	2	68	355	3.95	2	8	52	379	4.42	2	6	50
Parietal	Angular Gyrus	LH	858	6.91	-38	-54	36	309	5.04	-38	-56	38	366	5.85	-36	-56	36					
		RH	481	5.49	34	-54	36	250	4	34	-54	36	321	5.6	44	-48	46					
	Central Opercular Cortex	LH	313	4.6	-46	2	12	29	2.97	-54	-22	22	507	4.6	-50	8	-2	214	4.33	-54	-20	20
			112	4.36	-56	-22	20											28	2.6	-40	-6	12

	RH											57	3.38	42	0	10					
Cingulate Gyrus posterior	LH	1146	5.84	-4	-28	26	29	2.95	0	-36	2	623	4.49	-6	-28	24					
												30	3.75	-20	-28	-6					
	RH	343	5.56	2	-30	26						199	3.73	2	-30	26					
		44	3.57	2	-40	2															
Parietal Operculum Cortex	LH	163	4.51	-44	-38	32						283	4.77	-44	-38	32	240	4.28	-50	-24	18
Postcentral Gyrus	LH	2101	5.19	-36	-38	40	1752	4.87	-42	-34	44	2713	6.43	-40	-36	42	2304	4.6	-38	-28	64
		376	3.92	-2	-40	70	43	2.77	-2	-40	72										
	RH	469	4.6	54	-16	30	143	3.37	44	-26	38	992	5.03	44	-28	42	61	3.33	42	-36	52
		29	3.24	2	-38	70															
Precuneous Cortex	LH	2018	6.48	-8	-70	38	793	4.19	-4	-74	40	1558	5.75	-8	-70	38					
		44	3.63	-2	-52	-2															
	RH	795	4.79	2	-74	40	46	3.21	2	-74	42	508	4.41	4	-70	44					
Superior Parietal Lobule	LH	1673	6.7	-34	-56	42	1019	5.13	-42	-48	50	1430	6.4	-42	-44	46	1038	4.79	-28	-58	48
	RH	642	5.37	34	-56	42	306	4.01	32	-56	50	854	5.54	38	-48	46	377	3.75	42	-38	52
Supramarginal Gyrus anterior	LH	936	6.4	-50	-40	48	634	5.17	-50	-38	44	883	6.59	-42	-36	42	839	4.74	-48	-36	46

		RH	312	5.02	48	-34	40	99	3.46	44	-28	40	309	4.97	44	-30	42	55	3.11	46	-32	46
	Supramarginal Gyrus posterior	LH	888	6.64	-48	-44	50	478	5.85	-46	-46	48	603	6.52	-44	-42	46	226	4.26	-46	-40	48
													45	3.3	-48	-50	12	81	3.63	-56	-42	18
		RH	584	5.88	44	-40	40	295	3.73	46	-40	40	495	5.51	44	-46	44	106	3.82	44	-38	52
Temporal	Heschls Gyrus	LH											38	4.04	-46	-24	14					
	Inferior Temporal Gyrus posterior	LH	417	4.8	-46	-44	-20	39	3.48	-46	-44	-18	146	4.39	-46	-44	-20	25	3.03	-46	-42	-20
		RH	67	3.42	60	-38	-16															
	Inferior Temporal Gyrus temporooccipital	LH	1016	6.28	-46	-62	-18	500	4.66	-50	-62	-24	646	6.21	-42	-62	-10	437	5.88	-46	-62	-18
		RH	187	4.62	48	-62	-20	32	3.44	52	-58	-28	132	4.61	46	-60	-16	225	4.31	46	-60	-16
			179	3.75	58	-40	-16															
	Lingual Gyrus	LH	763	5.5	-12	-90	-12	301	5.18	-12	-90	-12	389	5.18	-12	-90	-8	472	5.74	-12	-90	-12
		RH	814	6.85	8	-78	-26	353	5.9	14	-88	-6	502	6.45	16	-88	-8	345	5.82	16	-88	-8
			30	3.23	6	-38	-4															
	Middle Temporal Gyrus posterior	LH	818	4.53	-66	-34	-6						444	3.82	-68	-38	-4					
		RH	198	3.54	68	-38	-14															

	Middle Temporal Gyrus temporooccipital	LH	1069	4.72	-56	-52	6						538	3.66	-56	-54	4	50	3.96	-44	-60	0
		RH	86	3.71	70	-40	-16															
	Planum Temporale	LH																29	3.66	-54	-42	18
	Superior Temporal Gyrus posterior	LH	123	4.07	-62	-38	2						38	3.28	-66	-34	0					
	Temporal Fusiform Cortex posterior	LH	363	6.76	-36	-50	-28	131	4.38	-40	-52	-28	361	5.69	-36	-50	-28	257	5.75	-36	-48	-28
		RH	235	5.22	36	-42	-30	83	4.4	36	-42	-30	166	4.97	34	-40	-30	70	4.05	36	-42	-30
	Temporal Occipital Fusiform Cortex	LH	729	6.56	-36	-50	-26	544	5	-42	-58	-26	655	6.3	-42	-60	-16	677	6.23	-42	-62	-22
		RH	763	6.52	34	-50	-26	449	5.34	40	-58	-26	665	5.82	36	-52	-26	655	5.45	40	-60	-22
	Temporal Pole	LH	645	5.02	-54	14	-6	92	3.79	-54	14	-4	472	5.68	-52	12	-4	277	3.99	-54	12	-10
													170	3.15	-32	22	-36					
		RH	181	3.93	58	14	-8						138	3.92	52	16	-4					
Occipital	Cuneal Cortex	LH	72	4.2	-4	-76	30															
		RH	43	3.85	2	-88	46															
	Intracalcarine Cortex	LH	175	4.4	-18	-88	2	49	4	-18	-88	2	66	4.35	-18	-88	2	268	4.82	-18	-88	2
		RH	87	2.93	10	-72	4	47	4.38	12	-88	0	57	4.51	12	-88	0	185	4.21	12	-88	0

			62	3.91	12	-88	0																			
	Lateral Occipital Cortex inferior	LH	2189	7.21	-36	-86	-8	1576	6.2	-30	-86	-8	1996	6.84	-30	-86	-10	2058	7.21	-30	-84	-8
		RH	2020	7.09	32	-86	-14	1690	6.15	30	-86	-6	1960	6.46	32	-84	-6	1839	6.33	38	-86	-8
	Lateral Occipital Cortex superior	LH	3984	6.62	-28	-68	44	1865	4.71	-24	-62	52	2566	5.41	-28	-66	40	656	4.93	-28	-60	50
								63	3.81	-26	-88	8	181	4.4	-30	-88	10	161	4.4	-22	-90	8
		RH	1042	5.4	32	-58	42	222	3.71	30	-58	50	1048	5.17	34	-58	42	280	4.94	26	-88	6
			225	4.93	26	-88	6	132	4.81	26	-88	6	173	5.1	26	-88	6	127	3.29	30	-58	52
			37	3.36	4	-82	56															
	Occipital Fusiform Gyrus	LH	1544	6.81	-30	-86	-18	1296	6.41	-20	-90	-10	1422	6.69	-28	-86	-10	1504	7.08	-28	-84	-8
		RH	1412	7.48	24	-88	-12	1241	5.96	30	-84	-6	1373	6.47	20	-88	-10	1382	6.3	22	-88	-12
	Occipital Pole	LH	2006	6.92	-16	-94	-8	1923	6.53	-18	-92	-10	1925	7.31	-26	-92	-2	2103	7.06	-18	-94	-4
		RH	2078	6.83	24	-90	-6	1773	6.88	14	-92	-8	1733	6.28	22	-94	0	1834	6.9	24	-92	-4
Medial	Amygdala	LH	159	4.22	-20	-10	-12						129	3.44	-28	-8	-14					
		RH	55	3.45	14	0	-16															
	Hippocampus	LH	279	4.25	-22	-26	-8						118	3.83	-22	-30	-6					
		RH	61	3.78	20	-28	-8															

	Insular Cortex	LH	663	5.77	-34	22	0	187	4.45	-32	22	-4	688	5.31	-32	22	-2	204	3.45	-32	-6	12
		RH	481	5.22	32	20	0						468	4.75	40	18	-2					
			34	3.04	38	0	10															
	Parahippocampal Gyrus anterior	LH	57	3.39	-16	-6	-18						46	3.45	-22	0	-14					
	Parahippocampal Gyrus posterior	LH	441	4.28	-10	-32	-8						195	3.83	-24	-32	-6					
		RH	189	3.78	20	-28	-8						55	3.05	10	-30	-8					
Basal Ganglia	Caudate	LH	441	5.74	-14	2	14	301	4.75	-12	4	10	393	4.73	-10	-2	12					
		RH	352	5.15	10	10	2	97	3.26	12	8	10	282	4.1	12	10	8					
	Pallidum	LH	187	4.72	-12	4	-2	146	4.02	-14	6	0	125	3.84	-16	6	0	34	3.05	-24	-16	0
		RH	101	3.91	16	8	-2	102	3.86	16	2	-6	26	2.95	16	8	0					
	Putamen	LH	652	4.59	-16	10	-2	168	4.14	-16	10	-2	661	4.45	-26	4	-4	116	3.55	-26	-10	12
								48	3.86	-24	-8	10										
		RH	432	4.99	14	12	-6	40	3.72	16	6	-8	509	3.76	26	16	2					
Midbrain	Accumbens	LH	78	4.53	-8	16	-2	40	4.28	-10	16	-4	68	3.47	-8	16	-2					
		RH	72	4.87	12	12	-6	42	3.4	10	16	-4										
	Thalamus	LH	1388	5.31	-10	-2	10	663	4.76	-10	-2	10	1309	4.84	-10	-4	10	194	3.54	-10	-18	4

RH	765	5.1	2	-20	10	57	3.06	2	-6	10	690	4.39	2	-14	14
	37	3.71	22	-26	-6						46	3.1	22	-30	0

Table S4. Activation coordinates for contrasts between conditions during the test phase.

			Aloud-Silent					Aloud-Control					Control-Silent				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	35	2.75	-12	38	10										
	Frontal Orbital Cortex	LH	96	3.16	-28	18	-24										
			89	3.1	-30	34	-20										
	Frontal Pole	LH	257	3.15	-20	62	20										
			164	3.43	-26	40	-22										
		RH	72	3.42	12	66	-16										
	Inferior Frontal Gyrus pars opercularis	LH	258	3.86	-44	6	18						92	2.82	-38	10	28
	Middle Frontal Gyrus	LH											159	3.54	-34	0	42
	Paracingulate Gyrus	LH	44	3.49	-8	48	12										
	Precentral Gyrus	LH	145	3.86	-46	4	18						335	3.62	-34	-4	42
													89	3.67	-32	-26	46
		RH											30	2.7	32	-18	42
Parietal	Angular Gyrus	LH	404	3.46	-50	-56	32						109	3.34	-36	-54	38

		RH	30	2.67	52	-46	42	30	2.86	46	-48	44
	Central Opercular Cortex	LH	97	3.67	-44	4	14					
	Cingulate Gyrus posterior	LH	450	3.67	-6	-30	38	59	2.95	-10	-48	34
	Postcentral Gyrus	LH						417	3.91	-32	-30	46
		RH						234	3.55	50	-18	40
	Precuneous Cortex	LH	541	3.33	-12	-60	40	124	2.86	-12	-62	42
		RH	36	2.71	16	-64	46	28	2.84	14	-64	46
	Superior Parietal Lobule	LH						212	3.27	-36	-54	44
								26	2.74	-42	-42	54
		RH	44	3.01	34	-44	48	258	3.66	34	-46	50
	Supramarginal Gyrus anterior	LH	72	2.76	-42	-40	38	272	3.86	-48	-28	32
		RH						46	2.98	46	-30	44
	Supramarginal Gyrus posterior	LH	170	3.41	-40	-44	38					
		RH	88	2.94	34	-46	32	107	3.34	46	-38	46
Temporal	Inferior Temporal Gyrus posterior	LH	175	3.5	-52	-28	-16					
		RH	25	3.19	58	-36	-20					
	Inferior Temporal Gyrus temporooccipital	LH	135	3.69	-58	-58	-10					

				73	3.22	-54	-52	-26				
			RH	102	3.23	52	-42	-14				
	Lingual Gyrus		LH	60	2.79	0	-80	-22				
			RH	141	3.16	4	-80	-20				
				97	3.4	14	-60	-14				
	Middle Temporal Gyrus posterior		LH	540	3.7	-54	-28	-14				
			RH	98	3.17	72	-32	-14				
	Middle Temporal Gyrus temporooccipital		LH	631	3.83	-60	-58	-8				
			RH	55	3.06	68	-44	-14				
	Superior Temporal Gyrus posterior		LH	59	3.33	-60	-36	0				
	Temporal Occipital Fusiform Cortex		RH	237	3.37	36	-50	-20				
	Temporal Pole		LH	75	2.85	-36	24	-28				
				71	3.48	-50	10	-26				
Occipital	Lateral Occipital Cortex inferior		LH	117	3.12	-56	-80	-2				
				27	3.49	-56	-68	12				
			RH	85	3.58	26	-78	-28				
	Lateral Occipital Cortex superior		LH	1128	3.86	-34	-78	40	302	3.16	-40	-64	46

		RH	315	3.62	40	-68	34						
								84	2.81	22	-56	48	
								74	2.88	32	-68	36	
	Occipital Fusiform Gyrus	RH	211	3.45	20	-80	-26						
Medial	Insular Cortex	LH	80	3.65	-40	2	6						
	Parahippocampal Gyrus posterior	LH	75	3.32	-8	-34	-12						
		RH	33	3.07	12	-28	-10						

Table S5. Activation coordinates for aloud, silent control and sensorimotor control conditions relative to foil items during the test phase.

			Aloud-Foil					Silent-Foil					Control-Foil				
Lobe	**ROI**	**Hemi**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**	**Cluster size (mm3)**	**Max z**	**x**	**y**	**z**
Frontal	Cingulate Gyrus anterior	LH	296	3.8	-2	22	34						186	3.45	-4	24	32
													33	2.77	-10	-4	32
		RH	46	3.44	2	22	34										
	Frontal Operculum Cortex	LH	263	4.1	-46	18	2						106	3.3	-50	12	-4
	Frontal Orbital Cortex	LH	1064	4.11	-24	34	-24						365	3.8	-32	24	-4
													285	3.7	-24	24	-22
	Frontal Pole	LH	4692	5.54	-44	46	8	1184	4.13	-30	66	2	3193	4.85	-34	60	4
			103	3.61	-10	44	44										
		RH	319	3.57	24	70	-8						288	3.76	24	58	-6
			203	3.48	36	58	14										
			157	3.74	46	36	20										

Inferior Frontal Gyrus pars opercularis	LH	1101	4.45	-54	16	18	32	3.09	-34	18	24	778	4.09	-34	16	22
Inferior Frontal Gyrus pars triangularis	LH	880	4.15	-58	24	18	31	2.92	-48	28	22	728	4.07	-46	34	16
	RH	33	3.1	44	34	16										
Middle Frontal Gyrus	LH	2575	5.04	-44	20	32	785	4.49	-44	22	34	1793	4.65	-48	32	24
	RH	423	3.92	40	26	42						157	3.15	30	10	54
Paracingulate Gyrus	LH	759	4.35	-4	22	48						519	4.22	-4	28	38
	RH	204	3.94	2	22	40						68	3.31	2	22	42
Precentral Gyrus	LH	1489	4.36	-40	-2	50						1170	4.18	-30	-26	52
		86	3.13	-4	-14	76										
	RH											249	3.59	32	-10	50
Subcallosal Cortex	LH	59	4.54	-6	16	2										
	RH	30	3.47	8	16	-4										
Superior Frontal Gyrus	LH	1771	4.67	-18	24	54						245	3.97	-6	24	48
	RH	129	4.07	2	10	62						65	3.14	22	-10	54
Supplementary Motor Area	LH	373	4.59	-2	8	66										
	RH	126	4.06	2	8	62										

Parietal	Angular Gyrus	LH	819	5.53	-38	-58	44	286	4.26	-40	-56	40	359	5.2	-38	-56	38
		RH	415	4.48	48	-50	42						305	4.73	44	-48	46
	Central Opercular Cortex	LH	85	3.59	-44	4	14										
	Cingulate Gyrus posterior	LH	1050	5.29	-4	-28	30						560	4.5	-12	-48	34
		RH	235	4.84	2	-38	20						77	4.07	2	-38	20
			29	3.77	2	-40	2										
	Postcentral Gyrus	LH	517	3.74	-34	-26	56						847	4.4	-34	-30	46
													119	3.46	-10	-50	66
			202	3.57	-12	-48	66										
		RH	52	3.54	38	-30	38						375	4.22	52	-20	42
	Precuneous Cortex	LH	1942	6.06	-6	-70	36	467	4.69	-6	-70	34	1351	5.64	-6	-70	38
			31	4.2	-4	-52	12										
		RH	595	5.24	2	-70	36	34	4.11	2	-70	36	431	4.4	2	-72	38
			71	3.47	2	-74	64										
	Superior Parietal Lobule	LH	759	5.4	-34	-58	44	121	3.56	-34	-58	42	705	4.61	-34	-58	40
			77	3.24	-12	-50	68										
		RH	165	3.74	34	-48	38						294	4.67	38	-50	48

	Supramarginal Gyrus anterior	LH	309	4.19	-48	-40	38						405	4.24	-50	-40	48
		RH	129	3.97	46	-34	38						183	3.71	44	-28	40
	Supramarginal Gyrus posterior	LH	666	5.29	-46	-46	44	260	3.91	-48	-46	50	496	4.53	-42	-52	42
		RH	443	4.66	46	-40	38						424	4.48	46	-46	46
Temporal	Inferior Temporal Gyrus posterior	LH	258	4.46	-62	-32	-20						31	3	-56	-38	-16
	Inferior Temporal Gyrus temporooccipital	LH	671	4.36	-66	-46	-16										
		RH	42	3.38	54	-58	-28										
	Lingual Gyrus	LH	132	3.79	-2	-82	-16										
		RH	374	4.75	6	-78	-24						47	3.96	6	-78	-26
	Middle Temporal Gyrus posterior	LH	733	5.06	-64	-36	-8						354	4.01	-68	-40	-4
	Middle Temporal Gyrus temporooccipital	LH	694	4.95	-64	-46	-8						183	3.53	-58	-44	-8
	Superior Temporal Gyrus posterior	LH	59	4.04	-64	-34	0										
	Temporal Fusiform Cortex posterior	LH	25	2.98	-28	-38	-20										
		RH	80	3.37	28	-42	-26										
	Temporal Occipital Fusiform Cortex	LH	127	3.44	-48	-58	-24										

		RH	196	4.34	28	-46	-24						25	3.23	28	-46	-24
	Temporal Pole	LH	96	3.98	-36	24	-26						146	3.41	-54	20	-14
													93	3.34	-38	24	-26
			91	3.91	-56	18	-8										
Occipital	Cuneal Cortex	LH	50	3.84	-4	-76	30										
		RH	31	3.2	4	-92	40										
	Lateral Occipital Cortex inferior	LH	169	3.63	-64	-62	-12										
		RH	399	4.3	38	-74	-28						103	3.62	40	-72	-28
	Lateral Occipital Cortex superior	LH	3008	6.25	-32	-74	38	729	3.9	-38	-60	38	1791	5.32	-38	-66	46
								67	2.94	-10	-68	54					
		RH	388	4.1	40	-70	38						316	4.24	38	-60	40
													199	3.5	14	-70	56
			49	3.31	14	-72	44										
			31	3.39	4	-80	58										
	Occipital Fusiform Gyrus	RH	479	5.01	10	-82	-24						210	3.92	10	-74	-24
	Occipital Pole	RH	78	3.41	2	-92	36										
Medial	Hippocampus	LH	127	3.45	-22	-26	-12										

	Insular Cortex	LH	288	4.39	-36	20	-2					328	4.33	-32	20	-4
		RH	38	3	32	20	4					42	2.92	28	16	4
	Parahippocampal Gyrus posterior	LH	299	3.83	-2	-40	0									
		RH	67	3.16	12	-34	-4									
Basal Ganglia	Caudate	LH	401	5.26	-10	4	12	191	4.23	-10	6	6	295	4.68	-10	4	8
		RH	288	4.26	10	10	6						172	3.71	12	10	6
	Pallidum	LH	70	4.74	-12	4	0	31	3.94	-14	6	0	40	3.97	-14	4	2
	Putamen	LH	142	3.84	-16	8	-2						187	3.55	-16	8	-2
		RH	43	3.21	14	14	-6						50	2.92	22	14	4
Midbrain	Accumbens	LH	73	4.37	-8	8	-4						46	3.59	-6	10	-6
		RH	50	3.86	10	14	-4										
	Thalamus	LH	895	4.84	-2	-18	12	131	3.89	-8	-2	8	550	4.41	-10	-2	10
		RH	357	4.44	2	-18	12						300	3.73	2	-16	14

Table S6. Activation coordinates for the silent condition relative to the aloud condition during the study phase.

Lobe	ROI	Hemi	Cluster size (mm3)	Max z	x	y	z
Frontal	Middle Frontal Gyrus	LH	372	4.49	-44	24	36
Frontal	Middle Frontal Gyrus	LH	85	3.84	-42	14	46
Parietal	Angular Gyrus	LH	94	3.59	-46	-58	42
Occipital	Lateral Occipital Cortex superior	LH	328	3.96	-44	-68	50
Occipital	Lateral Occipital Cortex superior	RH	229	4.07	42	-70	50

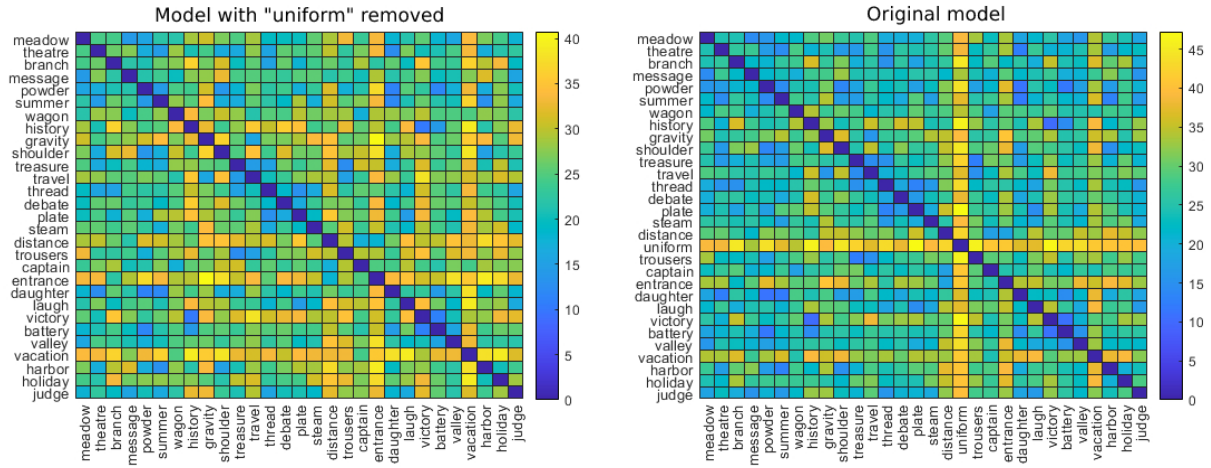


Figure S1. An example of a phonological model used for RSA represented as a heatmap that is symmetric about the diagonal. Colour scale reflects phonological edit distance and labels along the x and y axes represent study items included in the model. Left panel: a model used in the present study, right panel: the original model containing the extreme phonological outlier “uniform”.

2.12.4. Supplementary References

Favila, S. E., Lee, H., & Kuhl, B. A. (2020). Transforming the Concept of Memory Reactivation. Trends in Neurosciences, 43(12), 939–950. https://doi.org/10.1016/j.tins.2020.09.006

Freund, M. C., Bugg, J. M., & Braver, T. S. (2021). A Representational Similarity Analysis of Cognitive Control during Color-Word Stroop. Journal of Neuroscience, 41(35), 7388–7402. https://doi.org/10.1523/JNEUROSCI.2956-20.2021

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision Research, 41(10), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Hall, K. C, Allen, B., Fry, M., Johnson, K., Lo, R., Mackie, S., & McAuliffe, M. (2017). Phonological CorpusTools, Version 1.3. [Computer program]. Available from http://phonologicalcorpustools.github.io/CorpusTools/

Hertrich, I., Dietrich, S., Blum, C., & Ackermann, H. (2021). The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Frontiers in Human Neuroscience*,*15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.645209

Murphy, K. A., Jogia, J., & Talcott, J. B. (2019). On the neural basis of word reading: A meta-analysis of fMRI evidence using activation likelihood estimation. Journal of Neurolinguistics, 49, 71-83.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Frontiers in neuroinformatics, 10, 27.

Python Software Foundation. Python Language Reference, version 3.5.6. Available at http://www.python.org

Seghier, M. L. (2013). The Angular Gyrus. The Neuroscientist, 19(1), 43–61. https://doi.org/10.1177/1073858412440596

Vaden, K. I., Halpin, H. R., Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from http://www.iphod.com.

2.13. Transition To Chapter 3 (Bridging Section)

Chapter 3 comprises an under-review manuscript that is intended to build upon the findings of Chapter 2. Chapter 2 reported univariate analyses of fMRI data collected during a production effect paradigm; Chapter 3 builds on this work by applying multivariate analysis (RSA; see section 1.6.) to data from a simple fMRI experiment in which participants saw words onscreen and read each one aloud or silently. Chapter 3 is framed in terms of general reading processes (rather than the production effect specifically); however, as argued throughout Chapters 3 and 5, many of the findings therein are informative as to encoding processes which likely contribute to the production effect.

The analyses presented in Chapter 3 (as well as Chapter 4) involve computation of feature-wise Bayes factors in order to test hypotheses at the group level. As this method is relatively new to the field of fMRI (and therefore may be unfamiliar to readers who are versed in conventional approaches to group-level statistical testing), it is worth outlining the method and underlying reasoning here to provide some context for inferences presented in Chapters 3 and 4.

To date, most fMRI studies have relied on frequentist statistical testing to determine whether their data support a null hypothesis H_0 (e.g., "there is no difference between conditions A and B") or an alternative hypothesis H_1 ("there is a true difference between A and B"). Typically, researchers will compute a statistic (often a *p* value or *z* score) which determines statistical significance—that is, whether H_0 may be rejected, at which point one would accept H_1 by default. Whether statistical significance is achieved depends on whether the computed statistic falls above or below an a-priori threshold (e.g., $p < .05$ or $z > 2.3$). In practice, a given contrast will yield a *p* or *z* statistic at every voxel, and H_0 is accepted or rejected at the level of clusters. Clusters, in turn, are defined as groups of contiguous, statistically significant voxels which meet some other a-priori criteria (e.g., spatial or statistical extent, or a combination of the two; S. M. Smith & Nichols, 2009).

The frequentist approach imposes two major limitations on statistical inference. First, achieving statistical significance simply means that one may reject H_0 (and therefore accept H_1); *p* and *z* values do not speak to quantitative strength of evidence for either hypothesis. Second, this approach is susceptible to errors arising from multiple comparisons. This is because a *p* value reflects an estimate of the long-run frequency at which the observed result would be elicited if the experiment were run many times, assuming H_0 were true (stated differently: the probability that the observed result is a false positive). Accordingly, $p < .05$ is taken as an acceptable threshold for significance, because it places the probability of a false positive at less than 5%. Performing multiple tests, therefore, raises the likelihood that a false positive will occur (Jafari & Ansari-Pour, 2019). A common solution to this problem is to adjust one's criteria for statistical significance in a manner that is proportional to the number of tests being performed. In the context of fMRI, clustering methods (e.g., threshold-free cluster enhancement; S. M. Smith & Nichols, 2009) correct for multiple comparisons across voxels in a manner that does not excessively punish the data (considering that a single fMRI contrast will involve a single test at each voxel). However, multiple sets of tests (e.g., multiple between-condition contrasts) require additional correction to ensure that the actual chance of a false positive remains at the desired threshold (Alberton et al., 2020). This imposes harsh penalties in cases where many contrasts are performed (for example, the analyses presented in Chapter 3 entail 15 independent statistical contrasts). As argued by Dienes (2016), this framework punishes curiosity, because asking more questions (i.e., performing more contrasts) demands increasingly stringent statistical corrections.

Bayes factors offer an alternative to the traditional approach described above, and are becoming increasingly popular for group-level analyses in multivariate decoding studies (Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; Kaiser et al., 2018; Matheson et al., 2023; Proklova et al., 2019; Teichmann et al., 2022). Unlike *p* values, which are concerned only with the long-run probability of a single hypothesis (H_0), Bayes factors reflect the probability of the data under a model of one hypothesis versus another. For example, the Bayes factor BF_{10} is a ratio expressing

the likelihood of the observed data under H_1 versus H_0 (Dienes, 2014, 2016; Schmalz et al., 2021). A major advantage of Bayes factors is that they quantify the strength of evidence supporting either hypothesis. BF_{10} values range from 0 to ∞, wherein values < 1.0 express evidence in favour of H_0, and values > 1.0 express evidence for H_1. Within this framework, deviations from 1.0 reflect increasing strength of evidence, such that incrementally larger values above 1.0 reflect increasingly strong evidence for H_1. By convention, Bayes factors are typically assigned qualitative labels to express quantitative strength of evidence: BF_{10} values < 3.0 are considered to reflect "weak" evidence for H_1, values < 10.0 reflect "moderate" evidence, and values ≥ 10.0 reflect "strong" evidence (Dienes, 2014; M. D. Lee & Wagenmakers, 2014). Because BF_{10} reflects the probability of the data under H_1 versus H_0 (rather than the long-run probability of H_0) one can interpret it at face value without needing to account for repeated testing (Dienes, 2016; Teichmann et al., 2022). Stated differently, the relative probability of a given data point under either model does not change if additional tests are performed. This removes the necessity for punishing post-hoc corrections in instances where multiple questions are asked of the data.

In Chapters 3 and 4 I will evaluate evidence for H_1—that there is a true difference in decodability between aloud and silent reading, or that decodability in either condition is greater than zero—by performing independent Bayesian *t*-tests at each voxel (separately for each contrast / analysis). This is inspired by recent M/EEG work in which Bayesian *t*-tests were performed at each discrete time point (e.g., to detect above-chance classification accuracy; Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; see Teichmann et al., 2022 for a thorough description of this approach). I refer to this approach as feature-wise Bayes factor computation, because BF_{10} is computed at each feature (i.e., discrete point of measurement), which might be time points (M/EEG) or voxels (fMRI). Having computed BF_{10} at each feature, one can interpret the evidence for H_1 at face value (Teichmann et al., 2022). Stated differently, this approach enables one to compute easily interpretable whole-brain statistical maps for a given contrast, wherein the value at a given voxel (or, more intuitively, the

distribution of values over a large area) reflects the strength of evidence for or against one's hypothesis.

CHAPTER 3. DIFFERENTIAL WEIGHTING OF INFORMATION DURING ALOUD AND SILENT READING: EVIDENCE FROM REPRESENTATIONAL SIMILARITY ANALYSIS OF FMRI DATA

3.1. Publication Information

Bailey, L.M., Matheson, H.E., Fawcett, J.M., Bodner, G.E., & Newman, A.J. (under review). Differential weighting of information during aloud and silent reading: Evidence from representational similarity analysis of fMRI data. Manuscript ID: IMAG-24-0144[4]

3.2. Abstract

Single word reading depends on multiple types of information processing: readers must process low-level visual properties of the stimulus, form orthographic and phonological representations of the word, and retrieve semantic content from memory. Reading aloud introduces an additional type of processing wherein readers must execute an appropriate sequence of articulatory movements necessary to produce the word. To date, cognitive and neural differences between aloud and silent reading have mainly been ascribed to articulatory processes. However, it remains unclear whether articulatory information is used to discriminate unique words, at the neural level, during aloud reading. Moreover, very little work has investigated how other types of information processing might differ between the two tasks. The current work used representational similarity analysis (RSA) to interrogate fMRI data collected while participants read single words aloud or silently. RSA was implemented using a whole-brain searchlight procedure to characterise correspondence between neural data and each of five models representing a discrete type of information. Compared with reading silently, reading aloud elicited greater decodability of visual, phonological, semantic, and articulatory information. This occurred mainly in prefrontal and parietal areas implicated in speech production and cognitive control. By contrast, silent reading elicited greater decodability of orthographic information in right anterior temporal lobe. These results support an adaptive view of reading whereby information is weighted according to its task relevance, in a manner that best suits the reader's goals.

3.3. Statement Of Student Contributions To Manuscript

My contributions to this chapter included writing (original and subsequent drafts), study design, data collection, curation, analysis and interpretation, and creation of tables and figures.

[4] This chapter is a modified version of the manuscript submitted for peer review.

3.4. Introduction

Single word reading is an automatic and effortless process for most literate individuals. From a cognitive perspective, however, word reading may be considered as a series of computations whereby a printed stimulus is mapped to cognitively relevant representations. In order to characterise the neural underpinnings of word reading, it is useful (and, indeed, commonplace in cognitive neuroscience) to decompose this complex process into discrete types of information processing; while recognizing insights from recent language models which reveal complex interactions between different levels of representation (e.g., Caucheteux et al., 2023; Henningsen-Schomers & Pulvermüller, 2022). In this spirit, we consider five broadly-defined types of information that may be extracted from a printed word: visual, orthographic, phonological, semantic, and (in the case of reading aloud) articulatory. While neural processes concerning each type of information have been studied extensively, it remains unclear how, and to what extent they might differ between reading aloud and reading silently.

This question is particularly relevant in the context of embodied and grounded cognition; that is, the perspective that our internal cognitive states are influenced (if not determined) by the manner in which we receive input from and interact with our environment (Matheson & Barsalou, 2018). From this perspective, viewing cognitive processes (such as reading) as being fundamentally experience-dependent is essential for understanding their neural substrates. In many respects, reading aloud and reading silently entail fundamentally different experiential (and, therefore, cognitive) states. First and foremost, reading aloud entails additional motoric / articulatory processes required for speech production. In turn, the physical act of articulation elicits sensory experiences: the sensation of moving one's mouth and tongue, feeling one's larynx vibrate, and hearing one's own voice. Moreover, reading aloud has been shown to confer a reliable memory advantage (MacLeod et al., 2010), suggesting differences in cognitive processing beyond low-level motoric and sensorimotor mechanisms. This *production effect* is often attributed to distinctive sensory experiences brought about through articulation, though other mechanisms have been proposed. For example, there is evidence that words read

aloud benefit from more elaborate semantic processing (Fawcett et al., 2022), which may facilitate encoding and retrieval. All of this is to say that differences between aloud and silent reading extend beyond the mere physical act of articulation, and entail extensive cognitive differences. Although some work has investigated this possibility from a cognitive-behavioural perspective, it remains unclear how such differences might manifest at the neural level.

This question may be addressed with neuroimaging. In particular, representational similarity analysis (RSA) is a technique that allows us to decode different types of information (defined by formal hypothesis models) in neural patterns. The purpose of the present study was to investigate the presence of information in neural patterns associated with different types of information present in printed words, by applying RSA to functional magnetic resonance imaging (fMRI) data acquired during aloud and silent reading.

Below we provide a brief overview of the functional neuroanatomy of single word reading, emphasising cognitive and neural processes associated with each type of information described above. This is intended to provide some context for RSA literature on single-word reading and, ultimately, the design of the current study.

It is worth noting that most of the work discussed in the following section is based on experiments where Latin alphabet stimuli were read by speakers of Latin alphabet languages—predominantly English, but also others such as German (e.g., Pleisch et al., 2019), Italian (e.g., Liuzzi et al., 2020), and Finnish (e.g., Tarkiainen et al., 1999). In addition, dominant models of word reading (e.g., Coltheart et al., 2001; Seidenberg & McClelland, 1989) are predicated on rules of written English. Therefore, the findings discussed below may not be fully generalizable to speakers of languages using other scripts. That being said, fMRI studies exploring phonological analysis (C.-Y. Lee et al., 2004) and vocal production (Qu et al., 2022) in Chinese have revealed similar results to comparable studies in English, indicating consistency in the neural correlates of these processes across different scripts. Neural correlates of other components of words

reading, therefore, might also be consistent across different scripts. However, given that most work to date has focused on Latin alphabet speakers, we cannot assume that all of the following generalizes beyond those languages.

3.4.1. Functional Neuroanatomy Of Single Word Reading

When presented with a printed or written word, an individual must process its low-level visual properties—shape and orientation of the constituent letter strokes, size, colour, etc. At the neural level, these perceptual processes are largely governed by primary and associative visual cortices housed in occipital cortex (Cornelissen et al., 2009; Gramfort et al., 2012; Hauk et al., 2012; Tarkiainen et al., 1999). Distinct from this low-level perceptual processing is orthographic processing—the recognition of multi-character strings as *visual word-forms*. Visual word-forms may be described as perceptually invariant mental representations of words, irrespective of size, colour, font, or position in the visual field (Warrington & Shallice, 1980). With respect to the neural correlates of orthographic processing, much emphasis has been placed on the visual word form area (VWFA), located in the temperooccipital portion of the left fusiform gyrus. Neuroimaging work has demonstrated that this area responds preferentially to both real words and orthographically regular pseudowords compared to irregular consonant-string pseudowords (Cohen et al., 2002; Petersen et al., 1990; Polk & Farah, 2002), and also to real words and consonant strings compared to false font or unknown character strings (Baker et al., 2007; Brem et al., 2010; Carreiras et al., 2014; Pleisch et al., 2019). These findings indicate that VWFA is sensitive to letter strings generally, relative to perceptually similar non-letter visual objects, and also to orthographic regularity. In other words, it appears specialised for detecting visual stimulus properties that adhere to learned orthographic rules (i.e., statistics of written language), thus enabling recognition of familiar visual word forms. Moreover, studies on individuals suffering from pure alexia (selective impairment of the ability to recognise words rapidly and automatically, while identification of individual letters is preserved; McCandliss et al., 2003) often report

damage to VWFA (Pflugshaupt et al., 2009; Turkeltaub et al., 2014), suggesting a causal role of this area in orthographic processing.

During reading, orthographic information is associated to an auditory phonological code—that is, a mental representation of the constituent sounds that make up the spoken form(s) of the word they are reading (Leinenger, 2014). The nature of phonological processing in reading has historically been contentious, with opposing connectionist (e.g., Seidenberg & McClelland, 1989) and dual-route (e.g., Coltheart et al., 2001) models describing different mechanisms (see Seidenberg et al., 2022 for review and discussion). Here we adopt the connectionist principle of weighted spelling-to-sound mapping (Seidenberg, 2005)—which has been integrated into more recent connectionist-dual-process (CDP) models of word reading, and whose purpose is to bridge the gap between the two accounts (e.g., the CPP++; Perry et al., 2010). From this perspective, phonological forms are computed based on weighted connections between orthographic units and sounds; the weightings themselves are determined by the consistency of spelling-to-sound mappings within the language in question[5]. Words with inconsistent mappings may be considered to impose greater demands on the spelling-to-sound conversion system; hence, structures that are sensitive to spelling-sound consistency likely play a role in this process. A number of fMRI studies have revealed preferential activation of the left inferior frontal gyrus, and neighbouring structures such as anterior insula and anterior cingulate, when participants see words with inconsistent spelling-to-sound mappings relative to those with consistent mappings, implicating these structures in spelling-to-sound mapping (Bolger et al., 2008; Fiez et al., 1999; Fiez & Petersen, 1998; C.-Y. Lee et al., 2004).

Readers also experience semantic processing—that is, rapid and automatic retrieval of the meaning of the word they are reading. Based on a meta-analysis of 180

[5] For example, the vowel sound in *wave, cave, save,* etc. has high spelling-sound consistency in English, because *-ave* usually produces the eɪ vowel. By contrast, the vowel sound in *have* has low consistency because it does not obey this conventional spelling-sound mapping.

fMRI studies (Binder et al., 2009), Binder and Desai (2011) proposed a left-lateralized neurobiological model of semantic processing comprising dorsomedial and inferior prefrontal, inferior parietal, and inferior and ventral temporal cortices. Some of these same areas (particularly the inferior frontal gyrus, inferior parietal lobule, and inferior and superior temporal gyri) have also been identified by other meta-analyses of semantic processing (Rodd et al., 2015; Vigneau et al., 2006). In line with these meta-analyses, more recent evidence from RSA of fMRI data (described in more detail in 1.3) has supported the view that representation of semantic knowledge is distributed across bilateral prefrontal, parietal, and ventral temporal areas (Carota et al., 2021; Devereux et al., 2013; Liuzzi et al., 2020; Nastase et al., 2017).

In the case of reading aloud, an individual must engage additional articulatory (motor) processes, and will experience both proprioceptive feedback from moving one's mouth and tongue, and acoustic stimulation associated with hearing the sound of one's own voice. Indeed, a number of fMRI studies have reported that, compared to silent reading, reading aloud is associated with activation of auditory and sensorimotor cortices (Bailey et al., 2021; Dietz et al., 2005; Qu et al., 2022). Bailey and colleagues attributed their findings to motoric and sensory experiences involved in articulation, while Dietz and colleagues emphasised increased phonological processing demands in the context of reading aloud. In addition, a meta-analysis by Murphy et al. (2019) identified a single cluster in left STG that responded preferentially to aloud compared to silent reading. Although the authors did not offer an interpretation of this finding, one might take left STG activation to reflect auditory processing while reading aloud, given that STG has been implicated in both basic auditory processing, and speech production specifically (Hickok & Poeppel, 2007; Scott et al., 2000).

3.4.2 Investigating Aloud And Silent Reading With RSA

Recent years have seen growing popularity of multivariate analysis methods for fMRI, which are all broadly concerned with the distributed patterns of activation elicited by stimuli. One such approach is RSA (Kriegeskorte et al., 2008), which allows us to

characterise the informational content of these distributed activation patterns. Rather than making inferences based on contrasts between experimental conditions (as in univariate studies described above), RSA permits direct comparison between neural activation patterns and explicit hypothesis models, which often characterise specific stimulus information. Formally, this method quantifies the representational geometry of stimulus-specific activation patterns—that is, the geometric relationships between stimuli in high-dimensional space—as a *representational dissimilarity matrix* (RDM). This neural RDM may then be compared to one or more researcher-defined RDMs (hypothesis models) derived from quantitative measures of a given stimulus property. Within this framework, we can decode (i.e., detect) specific types of stimulus information (visual, orthographic, semantic, etc.) in a given patch of cortex, by demonstrating statistical dependence between a neural RDM (extracted from patterns in that patch) and a relevant hypothesis model (Kriegeskorte & Kievit, 2013).

Some prior RSA work has investigated decodability of different types of information during single word reading[6]. Much of this work has taken a region-of-interest (ROI)-based approach to examine patterns in ventral temperooccipital cortex, which is largely specialised for visual object processing, and notably includes the VWFA. Studies in which participants see visually presented words or logographic characters often report that phonological and orthographic information can be decoded in the left fusiform gyrus (Fischer-Baum et al., 2017; Qu et al., 2022; Zhao et al., 2017), while one study also decoded low-level visual information in its right-hemisphere homologue (Fischer-Baum et al., 2017). Semantic information may also be decoded in ventral temperoccipital cortices (Fischer-Baum et al., 2017; Wang et al., 2018), consistent with a larger body of work investigating semantic representation of images (as opposed to words) in this area (e.g., Devereux et al., 2013; Liuzzi et al., 2020; Nastase et al., 2017). Beyond ventral temporal cortex, phonological information has been decoded in left prefrontal and

[6] As with the univariate literature surrounding single word reading, much of the data has come from English-speaking participants reading English stimuli. That being said, a number of RSA studies have been conducted in Chinese (Li et al., 2022; Qu et al., 2022; Wang et al., 2018; W. Zhang et al., 2020; Zhao et al., 2017), which has strengthened the generalizability of this literature.

inferior parietal cortices (Fischer-Baum et al., 2018; Li et al., 2022), as well as middle and superior temporal regions (Li et al., 2022). Moreover, Zhang et al. (2020) had participants either read aloud, mouth, or imagine reading aloud simple consonant-vowel syllables, and reported that articulatory and acoustic information was decodable across a range of frontal, temporal, and parietal ROIs previously implicated in speech production and sensation. Taken together, these studies indicate that each type of information that we consider relevant to word reading may be decoded in multiple brain areas during single word reading.

Most of the studies described above focussed on aloud or silent reading tasks independently; a critical question, however, is whether the decodability of various types of information might *differ* between these two tasks. This question is particularly relevant to research on neural correlates of the production effect, which is defined as a contrast between aloud and silent reading. Moreover, from an embodied and grounded cognition perspective, this question is important for assessing the degree to which the neural correlates of reading are experience-dependent.

Decoding-based differences between aloud and silent reading would be consistent with prior RSA work showing that semantic decodability in particular is often task- or experience-dependent (e.g., when performing tasks that emphasise different semantic features of presented stimuli; Meersmans et al., 2022; Nastase et al., 2017; Wang et al., 2018). It is possible that decodability of other types of information is similarly variable when comparing aloud to silent reading. Indeed, there are many reasons to expect such differences, outlined as follows. First and perhaps most obviously, one might expect greater dependence on (and therefore increased decodability of) articulatory information during aloud versus silent reading, because speaking a word necessarily requires participants to execute an appropriate sequence of articulatory movements. In this context, articulatory information should be decodable in primary motor and/or premotor cortices (i.e., precentral gyrus and supplementary motor area). Indeed, previous fMRI work has shown that vocalising different phonemes elicits discriminable responses in the precentral gyrus (Pulvermüller et al., 2006); this area may therefore have

the capacity to represent whole words as linear combinations of articulatory features that are required to produce the constituent sounds.

Neural mechanisms distinguishing aloud and silent reading likely extend beyond the motoric components of articulation. Cognitive work surrounding the production effect—whereby words read aloud are more readily remembered compared to words read silently (MacLeod et al., 2010)—has emphasised the role of distinctive sensorimotor experiences during articulation. These include the sensation of moving one's mouth and tongue, feeling one's larynx vibrate, and experiencing auditory feedback from hearing one's own voice. At a higher cognitive level, reading aloud also entails planning and monitoring of speech output, and possibly sensory attenuation of efferent signals (corollary discharge) (Khalilian-Gourtani et al., 2022). At the neural level, we might expect such distinctiveness to be manifest as changes in decodability of articulatory and/or phonological information, perhaps in higher-level (i.e., non-sensorimotor) areas associated with speech planning and monitoring, or episodic encoding. Broadly speaking, the former processes have been linked to medial and lateral prefrontal cortices (Bourguignon, 2014; Hertrich et al., 2021), while episodic encoding is thought to depend largely on frontoparietal and medial temporal areas (e.g., H. Kim, 2011).

There is also evidence that reading aloud enhances semantic processing (Fawcett et al., 2022). We might therefore expect to see higher semantic decodability during aloud versus silent reading. It has also been suggested that people allocate more attention to words that they read aloud (Fawcett, 2013; Mama et al., 2018; Ozubko, Gopie, et al., 2012). While attention is not clearly defined in this context (particularly in terms of its relationship to specific types of information), we might broadly consider this term to reflect changes in one's cognitive state whereby words read aloud are assigned greater weighting (i.e., they are prioritised) during perceptual and goal-directed cognitive operations, compared to words read silently. This interpretation is consistent with prior conceptualizations of attention as the allocation of (limited) cognitive resources to a stimulus (P. A. MacDonald & MacLeod, 1998; Mama et al., 2018). Upon seeing a cue to read an upcoming word aloud, participants may experience an overall increased level of

arousal as they prepare to map upcoming visual/orthographic information onto an appropriate vocal response. Increased arousal may lead to greater cognitive "investment" in processing multiple types of information (and therefore increased decodability of each), recognising that successful word production depends on accurate mapping of low-level visual information to orthography, phonology, and articulatory features.

The predictions outlined above have received partial support from two studies examining task-dependent changes in decodability during single word reading. Qu et al. (2022) reported that, relative to silent reading, aloud reading elicited greater decodability of orthographic and phonological information in the left anterior fusiform gyrus. Moreover, Zhang et al. (2020) presented participants with consonant-vowel syllables and reported that reading aloud, mouthing, and imagined reading aloud elicited differential correlations with an articulatory model in a number of areas implicated in speech production and sensation. Notably, articulatory information was decodable in left angular and superior temporal gyri when reading aloud, but not in the other two conditions. These two studies provide proof-of-principle that different reading tasks may modulate decodability of phonological, orthographic, and articulatory information. However, a number of questions still remain. Qu et al.'s (2022) analysis was confined to the fusiform gyrus; therefore, it is unclear how phonological and orthographic information (or, indeed, other types of information) might be differentially represented in other brain areas. Moreover, Zhang et al. (2020) did not explicitly compare aloud reading with silent reading (rather, these authors examined three production tasks with variable degrees of enactment), and so it remains unclear how articulatory decodability might differ between the former conditions.

3.4.3. The Current Study

In the above section, we showed that different kinds of information—visual, orthographic, phonological, semantic, and articulatory—may be decoded during single word reading. There is some evidence for differential decodability of phonological and

orthographic information between aloud and silent reading, but whether similar effects are present outside occipitotemporal cortex, or for other types of information, remains unclear. As such, the goal of the current work was to establish how reading words aloud versus silently affects the decodability of visual, orthographic, phonological, semantic, and articulatory information throughout the whole brain. We conducted an fMRI experiment in which participants saw visually presented words (each repeated multiple times throughout the experiment) and were instructed to read each word either aloud or silently. We used a searchlight procedure to generate a neural dissimilarity matrix centred on each voxel throughout the brain[7]; we then compared the neural data from each searchlight to hypothesis models representing each of the five types of information discussed above.

Given that articulatory information is essential for speech production, we predicted that, relative to silent reading, reading words aloud would result in increased articulatory decodability in primary and associative motor areas. Articulatory (and possibly phonological) information may also be present in frontoparietal and/or medial temporal areas associated with speech planning/monitoring and episodic encoding, consistent with accounts of the production effect which emphasise sensorimotor experiences during encoding of words spoken aloud. From the view that reading aloud enhances semantic processing, we predicted increased semantic decodability relative to silent reading. While univariate studies have implicated multiple areas in semantic processing (see Section 3.4.1), changes in semantic decodability seems most likely in ventral temperooccipital cortex, as this area has been implicated by multiple RSA studies. Finally, increased attention to aloud words may lead to increases in decodability of multiple types of information. Given that ventral temperooccipital cortex (particularly the

[7] The searchlight method entails "scanning" a relatively large area (which might be the entire brain or a pre-defined ROI of any size) by parcellating it into a series of smaller searchlight areas, with each centred on a single voxel. An activation pattern for each searchlight area is extracted from its constituent voxels, meaning that one may construct a neural RDM (and, subsequently, compare that RDM to any number of hypothesis models) corresponding to every point in the brain/ROI (Kriegeskorte et al., 2008; Oosterhof et al., 2016).

left fusiform gyrus) has attracted much attention concerning visual, orthographic, and phonological decodability, it seems plausible that such attentional effects would be present in this area. All of these predictions would be evidenced by higher correlations between the relevant hypothesis model(s) and neural patterns elicited by aloud, compared to silent reading.

3.5. Methods

3.5.1. Subjects

Data were collected from 30 participants, aged 18-40 (*M* = 21.43, *SD* = 4.57), 21 female, three left-handed. All participants reported normal or corrected-to-normal vision, proficiency in English, no history of neurological illness or trauma, and no contraindications to MRI scanning. Handedness information was obtained using the Edinburgh Handedness Inventory (Oldfield, 1971). Participants were recruited through on-campus advertising at Dalhousie University, and received $30 CAD reimbursement and a digital image of their brain. All data collection took place at the IWK Health Centre in Halifax, NS. All procedures were approved by the IWK Research Ethics Board. Participants provided informed consent according to the Declaration of Helsinki.

Four participants were excluded from data analysis: one reported difficulty reading visually presented words while in the scanner; one reported that they did not follow task instructions properly (they silently mouthed words in the aloud condition, instead of vocalising the words as instructed); one disclosed that they did not fit inclusion criteria (neurological normality) after having completed the study; one withdrew prior to completing all parts of the experiment. Therefore, data from 26 participants (18 female, two left-handed) were included in our analyses.

3.5.2. Stimuli And Apparatus

Participants viewed all stimuli while lying supine in the MRI scanner; stimuli were presented using the VisualSystem *HD* stimulus presentation system (Nordic Neuro Lab,

Bergen, Norway). Stimuli were presented on an LCD screen positioned behind the scanner and viewed by participants via an angled mirror fixed to the MR head coil. During the experiment, participants made button-press responses using MR-compatible ResponseGrip handles (Nordic Neuro Lab, Bergen, Norway), one in each hand. Stimuli were presented using PsychoPy 2020.2.1 (Peirce et al., 2019). We selected a subset of 30 nouns from Bailey et al. (2021) which, in turn, were sourced from MacDonald and MacLeod (1998). Words were 6 to 10 characters in length, each with a frequency greater than 30 per million (Thorndike & Lorge, 1944). Our full word list is presented in Supplementary Table 3.1.

All words were presented at the centre of the screen in white lowercase Arial font against a dark grey background (RGB: 128, 128, 128). Response instructions for each word (see Procedure) were grayscale icons presented at the start of each trial in the centre of the screen—participants were instructed to speak the word aloud if they saw a mouth icon, or to read the word silently if they saw an eye icon. For our active baseline task (see Procedure), individual numbers 1-9 were presented in white Arial font at the centre of the screen against the same grey background as the words.

Hypothesis Models. We constructed a set of hypothesis models to be used for RSA reflecting visual, orthographic, phonological, semantic, and articulatory properties of the words presented in our experiment. Unique hypothesis models were generated for each participant and condition separately (because words were randomly allocated to conditions for each participant, see Section 3.5.3) in the Python environment using custom scripting and publicly available Python packages. Each hypothesis model comprised a 15 x 15 square dissimilarity matrix (DSM) containing zeros along the diagonal; off-diagonal cells contained pairwise dissimilarity values corresponding to all possible pairs of words within that condition, and matrices were symmetrical about the diagonal. To generate each model, we computed dissimilarity values according to measures that were theoretically relevant to the respective type of information / stimulus property being examined. Each of these measures is briefly outlined below, with more details

provided in Supplementary Materials. Hypothesis models from a representative subject and condition are shown in Figure 3.1.

We computed visual dissimilarity as the correlation distance (1 - Pearson correlation) between vectorized binary silhouette images of each word as it was presented in the experiment (as in Kriegeskorte et al., 2008). This procedure was implemented using the Pillow package (Umesh, 2012). We computed orthographic dissimilarity as the correlation distance between unconstrained open bigram feature vectors (similar to Fischer-Baum et al., 2017), using the wordkit package (Tulkens et al., 2018). For phonology, words were vectorised according to their syllable-wise spelling-to-sound consistency (Chee et al., 2020), and dissimilarity was computed as the Euclidean distance between vectors (we found that using Euclidean distances produced overall more normally distributed hypothesis models, compared with other metrics such as correlation or cosine distance). While we are not aware of prior RSA work employing this measure, we suggest that spelling-sound consistency provides an appropriate measure of phonological processing as conceptualised in the Introduction. For semantic models, we computed cosine distances (1 - cosine similarity) between word2vec representations of each word's semantic content (consistent with previous work; Carota et al., 2021; Guo et al., 2022; Liu et al., 2021; Tong et al., 2022; Wang et al., 2018). Word2vec representations were acquired from the publicly available glove-wiki-gigaword-300 model (https://github.com/RaRe-Technologies/gensim-data) using the gensim package (Řehůřek & Sojka, 2010). Finally, we computed articulatory dissimilarity as the feature-weighted phonological edit distance (Fontan et al., 2016) between words, normalized for word length (Beijering et al., 2008; Schepens et al., 2012) using the PhonologicalCorpusTools package (Hall et al., 2019). While this is ostensibly a measure of phonological dissimilarity, it depends on articulatory features necessary to produce each word (e.g., place of articulation, movements of the tongue, teeth, and lips, etc.). As such, we feel that it provides a suitable means for modelling articulatory information. Moreover, our analyses required that our five measures be independent from one another (i.e., minimally correlated); as such, we were motivated to select measures of

phonological and articulatory dissimilarity that were as different as possible from one another.

To ensure that our five measures were independent from one another, we computed pairwise correlations between (vectorised) exemplar models. Exemplar models contained all words in our stimulus set, and were computed for each measure. We reasoned that, because stimulus allocation to each condition was randomised for each participant, comparing exemplar models (containing all possible stimulus combinations) was the best way to approximate potential correlations between models used in our analysis. Correlations between our five exemplar models are displayed in Table 3.1. Accompanying parenthetical values in Table 3.1 show Bayes factors BF_{10} (described in Section 3.5.6) for each correlation, computed using the bayesFactor package (Krekelberg, 2022) in the MATLAB environment, with default JZS priors (Rouder et al., 2009). These comparisons revealed moderate evidence for a true correlation between the phonological and semantic models (r = 0.106, BF_{10} = 3.38); all other comparisons revealed equivocal evidence (BF_{10} < 2.0).

Table 3.1. Correlation matrix for the five dissimilarity measures used in this study. Values are correlation coefficients for each pair of measures; values in parentheses are Bayes factors indicating strength of evidence of a true (anti-)correlation.

	Visual	**Orthographic**	**Phonological**	**Semantic**	**Articulatory**
Visual		-0.037 (0.051)	-0.043 (0.057)	0.03 (0.046)	> 0.001 (0.038)
Orthographic			-0.028 (0.045)	0.106 (0.449)	0.134 (1.934)
Phonological				-0.143 (3.382)	0.118 (0.792)
Semantic					0.01 (0.039)
Articulatory					

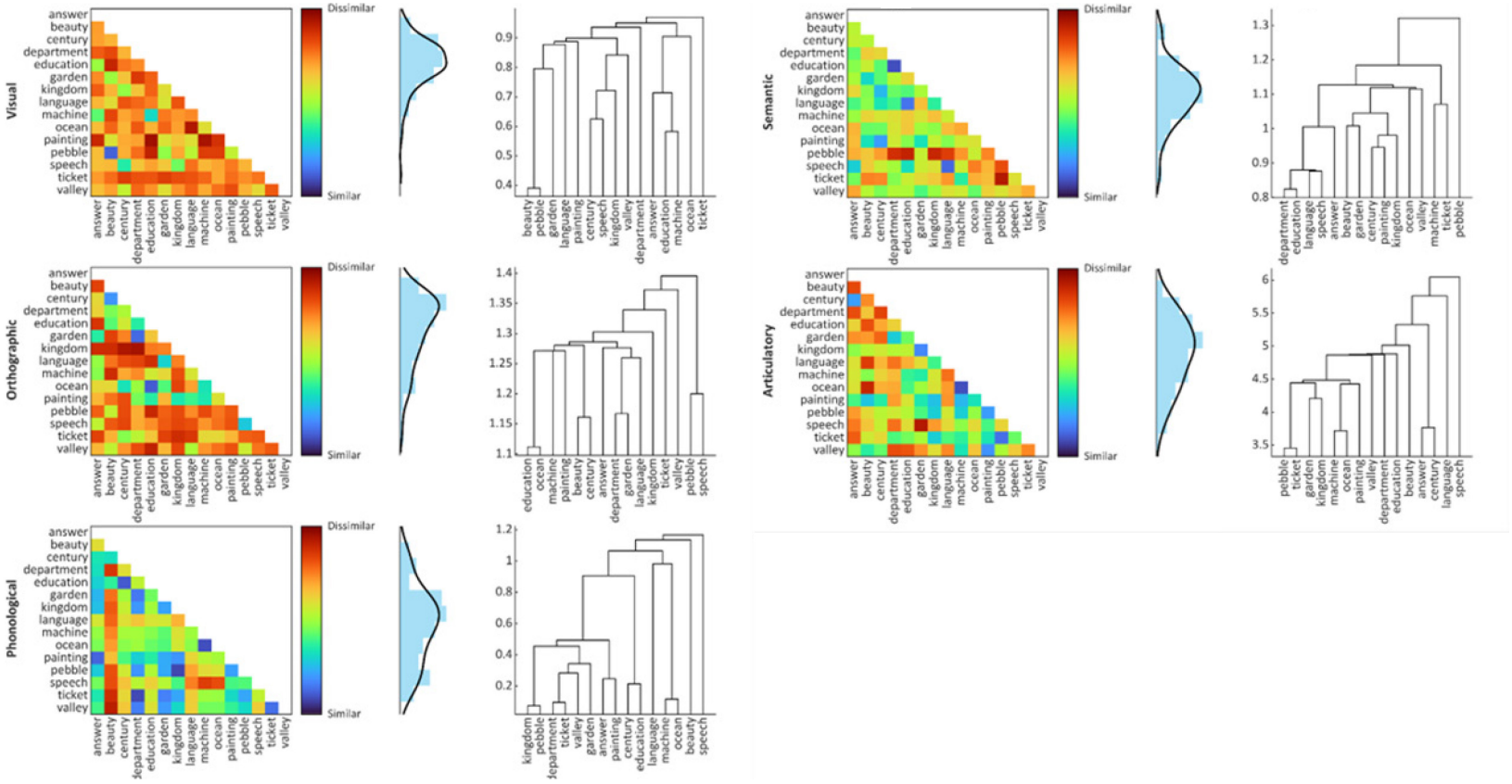


Figure 3.1. Hypothesis models (i.e., dissimilarity matrices) from a representative subject and condition. Values shown in heatmaps and histograms have been *z*-scored to reflect the data entered as regressors in each GLM (as described in Section 3.5.6). *3.5.3. Procedure*

After providing informed consent and completing MRI safety screening and the Edinburgh Handedness Inventory, participants completed a shortened version of the experiment described below (using a different stimulus list) on a laptop computer to familiarise them with the task. Participants were also asked to confirm (verbally) that they understood the task requirements before entering the scanner. Once participants were positioned in the MRI scanner, a brief scout scan determined head position. Following this, participants completed three functional scans as part of another study, followed by a structural scan. After the structural scan, participants completed four three-minute functional scans that comprised the data for the present study.

During each three-minute scan, participants performed a word reading task, in which the 30 words (Supplementary Table 3.1) were randomly allocated to one of two conditions (aloud or silent; 15 words in each condition) for each participant. Word-condition mappings were randomised between participants. Each trial of the word reading task began with a fixation cross ("+") presented for 500 ms to alert the participant that the trial was starting, followed by a cue presented for 1000 ms instructing participants to read the upcoming word either aloud (if they saw a mouth icon) or silently (if they saw an eye icon). Following the cue, a word from the corresponding condition was presented for 2500 ms. Each word appeared once in each functional run, and each of the four runs used a different random order of presentation.

On every trial, following word presentation, participants completed a short active baseline task: a randomly generated number between 1 and 9 was presented for 2000 ms, and participants were instructed to decide whether the number was odd or even, and to make an appropriate button-press response. During this baseline period, response mapping cues were presented as a reminder in the top corners of the screen (left-hand button press for odd, right-hand for even). Response mappings were the same for all participants. This active baseline served two purposes: (i) to lengthen SOA of each trial to 6 seconds, and therefore ensure reliable item-level estimation (Zeithamova et al., 2017), while (ii) ensuring that participants did not mind-wander between trials. We selected a simple odd-even judgement task because prior work has shown this task to be an effective

alternative to the traditional resting baseline in fMRI (Stark & Squire, 2001). A schematic of our trial structure is shown in Figure 3.2. Each functional run included 30 trials.

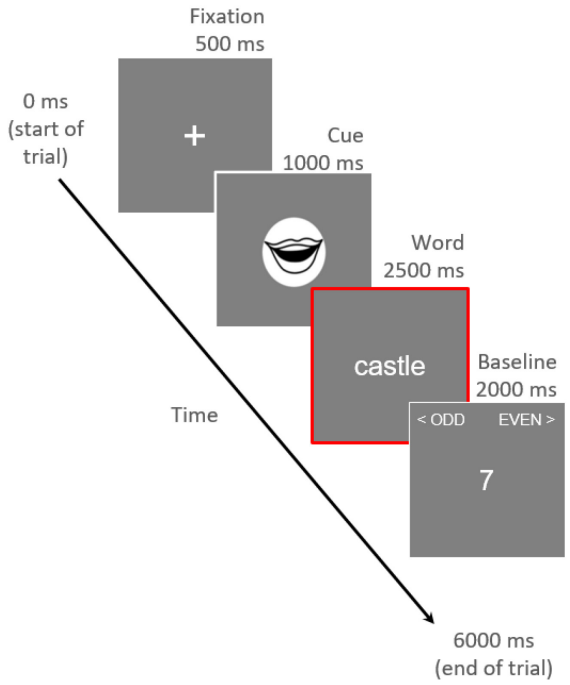


Figure 3.2. Schematic of an example trial. The red outline indicates the period modelled by the time series regressor for each trial (i.e., the temporal window from which activation patterns were estimated).

3.5.4. MRI Data Acquisition

MRI data was acquired on a 1.5 Tesla GE MRI system (GE Medical Systems, Waukesha, WI) equipped with a 19 channel head coil. Each participant completed a structural scan followed by four functional scans. As noted above, participants also completed three additional functional scans at the start of their session, but those functional scans are beyond the scope of the current work. For the structural scan, a T1-weighted anatomical image was obtained using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence, TI = 1134 ms, flip angle = 8°, NEX = 2, FOV = 224 mm, matrix = 224 x 224, resulting in an in-plane voxel resolution of 1 x 1 mm.

Functional scans used a gradient echo-planar pulse sequence, TR = 1800 ms, TE = 23 ms, FOV = 240 mm, flip angle = 90°. Images were obtained in 34 axial slices[8] (no gap, sequential acquisition) of thickness = 3.75 mm, matrix = 64 x 64, resulting in an in-plane voxel resolution of 3.75 x 3.75 mm. The FOV included full cortical coverage and partial cerebellar coverage. For each run we collected 100 functional volumes. Five additional volumes were acquired at the start of each run, but were discarded following acquisition.

3.5.5. fMRI Data Processing

All fMRI data processing was implemented with custom bash scripting, unless otherwise stated. To improve efficiency of our analysis pipeline, we used GNU Parallel (Tange, 2011) to perform many of the steps described below in parallel for multiple subjects/runs. The fMRI data were processed using functions from FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Preprocessing steps included non-brain removal with BET (S. M. Smith, 2002), grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor, high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s), and spatial smoothing (using a Gaussian kernel of FWHM 6 mm). We also performed motion correction using MCFLIRT (Jenkinson et al., 2002) to mitigate the impact of any potential movement during functional runs. We visually inspected motion correction output; our a-priori threshold for excessive motion between contiguous time points was 2 mm, however no runs exceeded this threshold, therefore no runs were removed due to excessive motion.

Following preprocessing, we used FEAT to estimate activation patterns for each trial using the least-squares-all (LSA) method (Mumford et al., 2014), described as follows. For each functional run, we constructed a single GLM that included one regressor for each trial in the run. Each regressor comprised a timeseries modelling the period during which

[8] We added 4 axial slices (total = 38) to the protocol for one participant to accommodate their entire cerebral cortex.

the word for that trial was presented (duration: 2500 ms), convolved with a gamma function (lag: 6 sec, sigma: 3 s) as a model of the haemodynamic response function (HRF). Additionally, we included parameters derived from motion correction as regressors of no interest. We defined 30 contrasts of interest, with each contrast comprising one regressor. This procedure generated four contrasts of parameter estimates (i.e., COPE images) for every unique item; one per trial. Because functional data from different runs are rarely perfectly aligned (due to small head movements between runs), we next aligned all the COPEs from each subject to a common native functional space. We used Advanced Normalization Tools (ANTs; http://stnava.github.io/ANTs/) to rigidly align COPEs from the second, third and fourth runs with the example functional volume (example_func in FSL) from the first fMRI run. We additionally masked the aligned data with a COPE image from the first run, thus removing non-overlapping voxels across runs.

Item-level pattern estimation (described as follows) was performed in the MATLAB environment using functions from the CoSMoMVPA toolbox (Oosterhof et al., 2016). We estimated item-level activity patterns for each unique item by averaging across COPEs from that item's four respective trials. We then subtracted the mean pattern (i.e., the mean value at each voxel across items) from item-level patterns in each condition separately, in order to remove the activation pattern common to all items in each condition (Walther et al., 2016)[9].

[9] While mean pattern subtraction is a contentious topic (Diedrichsen et al., 2011; Garrido et al., 2013), we argue that it is necessary in the case of our study. Shared mean activation has been shown to artificially inflate pairwise correlations between patterns (Walther et al., 2016); this in turn will add noise to correlation-based neural dissimilarity matrices. This has major implications for our study, the purpose of which was to compare decodability (that is, correspondence between neural DSMs and hypothesis models) between aloud and silent reading. Aloud reading reliably elicits more univariate activation than silent reading (Bailey et al., 2021; Dietz et al., 2005; Qu et al., 2022); in turn, item-level patterns for words read aloud likely share more common activation than do patterns for words read silently. As a result, we ought to see relatively greater inflation of pairwise correlations between patterns for words read aloud. Stated differently, failing to control for shared activation (within each condition) would likely result in systematically different degrees of noise contributing to decodability in aloud and silent reading, which would obfuscate any comparisons of decodability between these conditions.

3.5.6. Representational Similarity Analysis

Searchlight Analyses. We performed RSA using a whole-brain searchlight approach, separately for each subject and condition (aloud or silent reading), in each subject's native functional space. Searchlights were implemented in the MATLAB environment using functions from the CoSMoMVPA toolbox (Oosterhof et al., 2016). Within each searchlight area (spherical searchlight, radius = 3 voxels), we first generated a neural DSM comprising pairwise Pearson correlation distances (i.e., 1 minus the Pearson correlation) between item-level patterns for all items within that condition. Each neural DSM was vectorized and submitted to a GLM with five regressors, with each regressor comprising one of the five (vectorized) hypothesis models described above. Both the neural data and models were standardised using the z transform prior to estimating regression coefficients (Oosterhof et al., 2016). Each searchlight analysis generated five whole-brain statistical maps with a beta (β) value at each voxel; one map was generated for each hypothesis model. β values reflected correspondence between the neural DSM at that voxel (computed from patterns in the corresponding searchlight sphere) and the respective hypothesis model. A separate searchlight analysis was performed for each condition, such that each subject was associated with ten β maps in native functional space.

We next transformed each subject's searchlight output (i.e., whole-brain β maps) from native functional space to template MNI152 space using ANTs (http://stnava.github.io/ANTs/). We first computed a structural-to-EPI transformation matrix by rigidly aligning participants' high-resolution structural T1 image to their example functional volume from the first fMRI run. We next computed an MNI152-to-structural transformation matrix using linear affine, and then nonlinear methods (the

latter implemented using the "SyN" algorithm; Avants et al., 2008)[10]. We then applied the inverse of the structural-to-EPI and MNI152-to-structural transformation matrices to each subject's respective searchlight output, re-sliced to 2 mm isotropic resolution. Each subject was therefore associated with ten β maps in MNI152 space. Following this, we masked out the cerebellum (as defined by the Harvard-Oxford cortical atlas supplied by FSL) from all of the spatially normalised β maps; the motivation for this was that our fMRI scans included only partial cerebellar coverage.

Group-Level Analyses. For group-level analyses, we computed voxel-wise Bayes factors BF_{10} to evaluate our hypotheses. This is in contrast to conventional null-hypothesis statistical testing (e.g., permutation cluster statistics), which relies on *p* values to determine whether or not a null hypothesis may be rejected at each voxel. Unlike *p* values, Bayes factors quantify the strength of evidence in favour of a null or alternative hypothesis. BF_{10} is a ratio that expresses the likelihood of the data under the alternative hypothesis (H_1) relative to the null hypothesis (H_0) (Dienes, 2014), and may range between 0 and ∞. Within this framework, incrementally larger values > 1 indicate greater support for the data under H_1; incrementally lower values < 1 indicate greater support for the data under H_0. Recently a growing number of studies have adopted Bayes factors for group-level neuroimaging analyses, particularly in the context of MVPA (e.g., Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; Kaiser et al., 2018; Matheson et al., 2023; Moerel et al., 2022; Proklova et al., 2019; Teichmann et al., 2022). Teichmann et al. (2022) and Matheson et al. (2023) have argued in favour of Bayes factors over *p* values in MVPA research, primarily because Bayes factors are actually informative as to strength of evidence (as opposed to dichotomous

[10] To be clear: the structural-to-EPI and MNI152-to-structural transformation matrices were computed using the EPI and structural T1 as reference images respectively. While it is more common to compute these transformations in the opposite direction (i.e., EPI-to-structural and structural-to-MNI152, using structural T1 and MNI152 as reference images), we found that the inverse procedure yielded qualitatively better transformation of searchlight maps—that is, better alignment to the MNI152 template, based on visual inspection of registration output.

acceptance or rejection of H_0), and are not susceptible to multiple comparison problems (Dienes, 2016; Teichmann et al., 2022).

We computed Bayes factors using voxel-wise Bayes *t*-tests implemented with functions from the bayesFactor package (Krekelberg, 2022) in the MATLAB environment using default JZS priors (Rouder et al., 2009). We first tested whether each type of information was decodable in either condition. Here, we performed a right-tailed one-sample Bayes *t*-test at every voxel against the hypothesis that β was greater than zero; this procedure was repeated independently for each condition and type of information. This produced ten statistical maps of BF_{10} values (2 conditions x 5 information types); hereon in, we will refer to these as *within-condition* Bayes maps. Results from these analyses are shown in Supplementary Materials (Supplementary Figures 3.1 and 3.2, Supplementary Tables 3.2 and 3.3) but are otherwise not reported here; rather, they were used to constrain results from contrasts between aloud and silent reading, described as follows.

We next tested the hypothesis that there was a difference in decodability, for each type of information, between conditions. Here, we performed a two-tailed paired-samples Bayes *t*-test at every voxel against the hypothesis that β values differed between aloud and silent reading. This procedure was repeated independently for each type of information, generating five statistical maps of BF_{10} values; hereon in, *between-condition* Bayes maps.

As a convention, Bayes factors are often assigned qualitative labels to express the strength of evidence for H_1, whereby BF_{10} values < 3.0 are considered to provide "weak" or "anecdotal" evidence, values < 10.0 provide "moderate" evidence, and values >= 10.0 are considered to provide "strong" evidence (Dienes, 2014; M. D. Lee & Wagenmakers, 2014). We chose to only consider voxels where there was at least moderate evidence for any hypothesis (that is, decodability greater than 0, or differences between conditions); we feel that reporting weak evidence would not be particularly valuable, and moreover would distract from data points where there was relatively stronger support for our

hypotheses. As such, we thresholded all Bayes maps (both within- and between-conditions) at BF_{10} >= 3.0. To establish directionality of our results concerning differences between conditions, we first computed (for each type of information) two average contrast maps: an aloud > silent contrast representing the average searchlight results (across participants) from the aloud condition minus those of the silent condition; vice-versa for the silent > aloud contrast. These average contrast maps were used to mask the thresholded between-condition Bayes maps, thus generating one Bayes map representing aloud > silent, and one representing silent > aloud, for each type of information. Each contrast map was then masked by the within-condition map for its respective minuend condition; this ensured that the contrast maps only included voxels where information was actually decodable in the minuend condition (i.e., it served to exclude any between-condition differences driven entirely by negative values in the subtrahend condition).

We generated interpretable tables of results using FSL's cluster function to identify clusters of contiguous voxels, separately for each model and between-conditions contrast. We only considered clusters with a minimum spatial extent of 20 voxels, to avoid our results being skewed by noisy voxels. Tables report the spatial location (in MNI coordinates) of the centre of gravity (COG) for each cluster (ascertained using FSL's *cluster* function), as well as each cluster's mean (averaged over constituent voxels) and maximum BF_{10} value. Maximum BF_{10} values represent the strongest available evidence for a given hypothesis (e.g., for a difference between conditions) within a cluster, while we consider mean values to reflect evidential strength across the entire cluster. Tables also report anatomical labels for each cluster; these were identified from the Harvard-Oxford cortical and subcortical atlases using FSL's *atlasquery* function.

3.6. Results.

We performed RSA on fMRI data to investigate decodability of information across five types of information—visual, orthographic, phonological, semantic, and articulatory—during aloud and silent word reading. We used a whole-brain searchlight to

characterise, at every voxel, correspondence between the neural data and formal hypothesis models representing each type of information, for each condition separately. We then explored differences in decodability between the two conditions. Below, all references to "moderate" or "strong" evidence describe quantitative benchmarking of mean BF_{10} values that were computed for each cluster, based on conventional norms (Dienes, 2014; M. D. Lee & Wagenmakers, 2014).

3.6.1. Aloud Reading > Silent Reading

Visual, phonological, semantic, and articulatory—but not orthographic—information were all more decodable in the aloud reading condition compared to the silent reading condition; these results are displayed in Table 3.2 and Figure 3.3. For convenience, all references to decodability in this section describe differences between aloud and silent reading.

Visual Information. We identified a single cluster exhibiting moderate evidence for visual decodability. This cluster was situated in the occipital portion of the left fusiform gyrus (more precisely, on the fusiform bank of the posterior collateral sulcus; Lehman et al., 2016) and extended into lateral occipital cortex.

Phonological Information. Decodability of phonological information was notably more extensive than visual, semantic, or articulatory information, and was detected in bilateral frontal and parietal areas. In the frontal lobes, clusters were situated bilaterally on the ventrolateral / orbital surfaces of the frontal poles; notably, the cluster in the left frontal pole exhibited the strongest evidence for this contrast (mean BF_{10} = 172.96). In addition, a single cluster with strong evidence spanned prefrontal cortices in both hemispheres, encompassing dorsolateral (middle and superior frontal gyri, frontal pole) and dorsomedial (medial surfaces of the superior frontal gyrus and frontal pole) areas bilaterally. One additional cluster was situated on the lateral surface of the right precentral gyrus, though this cluster exhibited only moderate evidence. In the parietal lobes, clusters were situated in the precuneus bilaterally, though the right cluster

exhibited markedly stronger evidence (mean BF_{10} = 68.24 [right], 15.71 [left]) and was situated more posteriorly and ventrally, at the parieto-occipital boundary. Clusters were also present in the occipital lobe. On the lateral surface of the left hemisphere, one strong cluster was situated at the parieto-occipital boundary (mainly encompassing lateral occipital cortex) and, more posteriorly, a moderate cluster extended from the lateral surface of the occipital pole to posterior ventral occipital cortices. Finally, one cluster with moderate evidence was present in the occipital portion of the right fusiform gyrus.

Semantic Information. We identified two adjacent clusters in right prefrontal cortex (middle and superior frontal gyri), both of which exhibited moderate evidence for semantic decodability. The position of these clusters approximately corresponds to dorsal premotor cortex (Genon et al., 2017).

Articulatory Information. Articulatory information was decodable in medial frontal areas bilaterally, as well as in left premotor cortex. On the medial surface, two large clusters encompassed areas in both hemispheres. The anteriormost cluster encompassed anterior dorsomedial cortex (superior frontal and paracingulate gyri); the more posterior cluster was situated in the pre-supplementary motor area (pre-SMA). On the lateral surface, one cluster was situated in the posteriormost portion of the superior frontal gyrus; corresponding to the supplementary motor area proper (SMA). All of these clusters exhibited strong evidence.

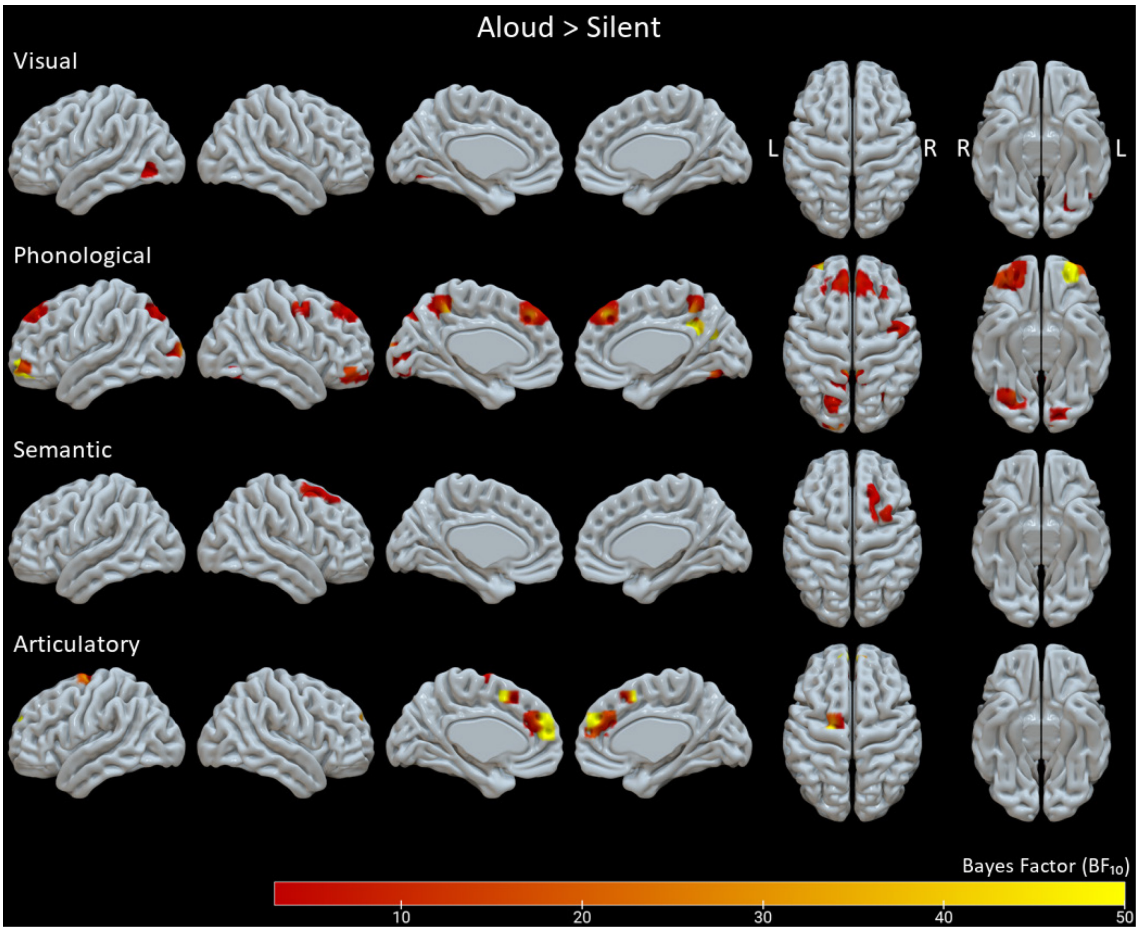


Figure 3.3. BF_{10} maps for the Aloud > Silent contrasts. Highlighted areas show evidence of greater decodability in the aloud reading condition relative to the silent reading condition. Surfaces from left to right show left lateral, right lateral, left medial, right medial, dorsal bilateral, and ventral bilateral views. *Table 3.2.* RSA resultSupplementary Table detailing clusters identified by the Aloud > Silent contrasts. Clusters show evidence of greater phonological, semantic, and articulatory decodability in the aloud reading condition relative to the silent reading condition.

Table 3.2. RSA results table detailing clusters identified by the Aloud > Silent contrasts. Clusters show evidence of greater phonological, semantic, and articulatory decodability in the aloud reading condition relative to the silent reading condition.

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Visual	1	58	9.41	29.9	-33	-69	-5	**L Lateral Occipital Cortex inferior (22)**
Phonological	1	118	172.96	882	-27	53	-8	**L Frontal Pole (118)**
	2	212	68.24	542	20	-55	29	**R Precuneous Cortex (141)**
								R Cingulate Gyrus posterior (29)
								R Cuneal Cortex (2)
								R Lateral Occipital Cortex superior (22)
								R Superior Parietal Lobule (2)
	3	116	15.71	57.3	-3	-48	54	**L Precuneous Cortex (83)**
								L Postcentral Gyrus (2)
								L Superior Parietal Lobule (1)
								R Precuneous Cortex (30)
	4	164	12.19	37.8	-18	-65	45	**L Lateral Occipital Cortex superior (136)**
								L Precuneous Cortex (24)
								L Superior Parietal Lobule (4)
	5	863	11.69	93.2	-1	39	46	**L Superior Frontal Gyrus (297)**
								L Frontal Pole (201)
								L Paracingulate Gyrus (2)
								R Frontal Pole (124)
								R Middle Frontal Gyrus (12)
								R Paracingulate Gyrus (5)
								R Superior Frontal Gyrus (220)
	6	140	10.41	23.2	29	48	-14	**R Frontal Pole (140)**
	7	190	9.02	35.1	-19	-90	5	**L Occipital Pole (90)**
								L Intracalcarine Cortex (23)

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								L Lateral Occipital Cortex inferior (15)
								L Lateral Occipital Cortex superior (13)
								L Lingual Gyrus (9)
								L Occipital Fusiform Gyrus (40)
	8	22	8.29	20.3	28	-72	-14	**R Occipital Fusiform Gyrus (22)**
	9	61	6.51	13.3	44	-2	50	**R Precentral Gyrus (60)**
								R Middle Frontal Gyrus (1)
Semantic	1	102	8.58	24.6	30	9	65	**R Middle Frontal Gyrus (41)**
								R Superior Frontal Gyrus (61)
	2	33	3.97	5.21	24	28	59	**R Superior Frontal Gyrus (32)**
								R Middle Frontal Gyrus (1)
Articulatory	1	368	36.71	294	0	49	25	**R Paracingulate Gyrus (115)**
								L Frontal Pole (5)
								L Paracingulate Gyrus (193)
								L Superior Frontal Gyrus (27)
								R Superior Frontal Gyrus (28)
	2	33	25.32	81.4	-19	-7	72	**L Superior Frontal Gyrus (33)**
	3	27	23.6	96	-3	15	52	**L Paracingulate Gyrus (12)**
								L Superior Frontal Gyrus (13)
								R Paracingulate Gyrus (1)
								R Superior Frontal Gyrus (1)

3.6.2. Silent Reading > Aloud Reading

Only orthographic information showed greater decodability in the silent reading condition compared to the aloud condition; these results are displayed in Table 3.3 and Figure 3.4. This contrast revealed two clusters with strong-to-moderate evidence (mean BF_{10} = 11.32, 8.93 respectively) in the right anterior temporal lobe. The stronger cluster spanned a relatively large area of anterior ventral temporal cortex, with the peak value (Max BF_{10} = 132) situated at the tip of the temporal pole, extending posteriorly into (anterior portions of) the fusiform and parahippocampal gyri. The second cluster was on the lateral surface, at the border between the temporal pole and superior temporal gyrus.

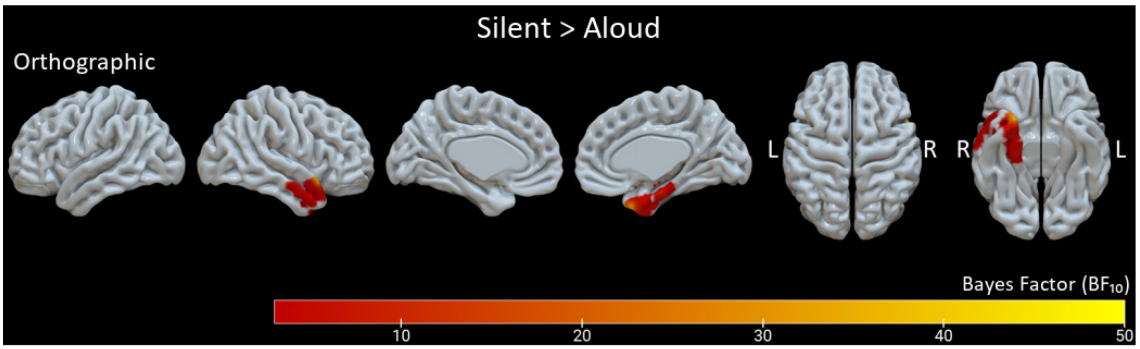


Figure 3.4. BF_{10} map for the Silent > Aloud contrast. Highlighted areas show evidence of greater decodability in the silent reading condition relative to the aloud reading condition. Surfaces from left to right show left lateral, right lateral, left medial, right medial, dorsal bilateral, and ventral bilateral views.

Table 3.3. RSA results detailing clusters identified by the Silent > Aloud contrast. Clusters show evidence of greater orthographic decodability in the silent reading condition relative to the aloud reading condition.

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Orthographic	1	357	11.32	85	27	7	-38	**R Temporal Pole (186)**
								R Inferior Temporal Gyrus anterior (4)
								R Parahippocampal Gyrus anterior (94)
								R Parahippocampal Gyrus posterior (3)
								R Temporal Fusiform Cortex anterior (70)
								Right Hippocampus (16)
	2	252	8.93	132	56	12	-21	**R Temporal Pole (179)**
								R Middle Temporal Gyrus anterior (58)
								R Middle Temporal Gyrus posterior (15)

3.7. Discussion

The current work aimed to characterise differences in the decodability of five types of stimulus-relevant information between aloud and silent single-word reading. We used RSA to compare fMRI data (acquired during each reading task) to formal hypothesis models representing visual, orthographic, phonological, semantic, and articulatory information. Our results revealed differential decodability for all types of information between the two tasks. Interestingly, and contrary to our initial predictions (see Introduction), we did not find evidence that decodability is unilaterally enhanced by aloud reading. Instead, visual, phonological, semantic, and articulatory information were all more decodable during aloud reading, while silent reading entailed greater decodability of orthographic information. We interpret these results to reflect differential weighting of information, depending on cognitive demands imposed by either task.

3.7.1 Decoding Information In Aloud Reading

We found moderate evidence that visual information was more decodable during aloud versus silent reading, in a cluster spanning the posterior portion of the left collateral sulcus and inferior lateral occipital cortex. In the context of reading, these posterior occipital areas have been associated with low-level orthographic analysis (Jobard et al., 2003; Levy et al., 2009); for example, they tend to respond preferentially to consonants versus false fonts (Thesen et al., 2012; Vinckier et al., 2007). We therefore suggest that our findings concerning visual information reflect greater attention (that is, allocation of more cognitive resources) to low-level perceptual properties of the visually presented words, consistent with the notion that participants allocate more attention to words spoken aloud (Fawcett, 2013; Mama et al., 2018; Ozubko, Gopie, et al., 2012).

Phonological and articulatory information were both decodable in dorsomedial prefrontal cortex (DMPFC). In addition, articulatory information was present in (pre-)supplementary motor areas (pre-SMA, SMA), while phonological information was present in dorsolateral prefrontal cortex (DLPFC) and the frontal poles. DMPFC, DLPFC, and (pre-)SMA are functionally connected to each other and to major language processing

centres in perisylvian cortex, and are all associated with planning and cognitive control of speech processes, amongst other functions (Bourguignon, 2014; Hertrich et al., 2016, 2021). Left DLPFC in particular is commonly associated with speech planning (see Hertrich et al., 2021 for review), while DMPFC has been linked to domain-general online response monitoring. For example, in the context of a Stroop task, DMPFC may monitor (and resolve) conflicting information arising on a trial-to-trial basis (Freund et al., 2021; C. Kim et al., 2013; A. W. MacDonald et al., 2000). With respect to its role in language production, a recent study linked DMPFC to learning arbitrary associations between visually presented stimuli and orofacial and vocal responses, as well as physically performing those responses (Loh et al., 2020). These authors also note that pre-SMA was recruited for learning speech-based vocal responses, but not non-speech vocalisations or orofacial movements, suggesting some degree of specialisation for articulatory processes. Pre-SMA has otherwise been associated with selecting and encoding complex motor sequences (Alario et al., 2006; Tremblay & Gracco, 2009, 2010), while SMA proper is thought to play a role in their implementation (Hertrich et al., 2016). As for the frontal pole, while its specific role in language production is less clear, it has been suggested that this region plays a domain-general role in goal-dependent feedback monitoring (Tsujimoto et al., 2011) and goal selection and maintenance (Fine & Hayden, 2021; Koechlin, 2011).

Given the above information, we suggest that the presence of phonological and articulatory information in the aforementioned areas reflects top-down planning and maintenance of vocal responses. To speak each word correctly, participants had to plan and coordinate an appropriate articulatory response comprising a unique sequence of movements of the mouth, jaw, and tongue. The presence of articulatory information in pre-SMA and SMA likely reflects preparation and implementation (respectively) of this complex motor sequence, while DMPFC may be involved in on-line monitoring to ensure that the produced responses match computed phonological information (see next paragraph). Thus while it is unsurprising that a vocal response should involve motor planning and monitoring, our results demonstrate that *stimulus-specific* articulatory

information can be decoded in these areas. As such, it is accurate to say that responses in SMA, pre-SMA, and DMPFC reflect unique articulatory properties of specific words.

With respect to phonological information (which in our study concerned spelling-to-sound consistency), executive control processes may be recruited to ensure that visually presented words are converted to appropriate phonological codes for the purpose of reading aloud. These control processes likely include speech planning (DLPFC), online monitoring during speech production (DMPFC) and goal-related cognition (frontal poles). With respect to the latter point, phonological information might be relevant to goal maintenance (e.g., updating one's current goal—to read the presented visual word aloud—to incorporate the desired sounds) or goal-related feedback (e.g., registering that the perceived visual word was produced as planned). As an aside, our interpretations here may seem at odds with the view that whole phonological forms of familiar/high frequency words (such as our stimuli) are retrieved automatically, without the need for explicit spelling-to-sound mapping (Coltheart et al., 2001). Direct phonological retrieval may have occurred in our study, but been supplemented by additional spelling-to-sound conversion, perhaps as an internal "check" to ensure that the appropriate sounds were retrieved. This suggestion is consistent with the view that the degree to which readers engage in explicit spelling-to-sound conversion is dependent upon their goals and/or task demands (Harm & Seidenberg, 2004). In our case, checks on phonological output may be driven by participants' goal to read the words aloud.

Phonological information was also present in the precuneus. Notably, the area identified by our analysis appears to correspond to part of the paracingulate network (Dadario & Sughrue, 2023). This network has been proposed to support integration of external input and prior knowledge to guide goal-directed behaviour, via its functional connections to other areas. These other areas notably include the insula and anterior cingulate (both implicated in spelling-to-sound conversion; see Introduction), medial frontal areas (possibly involved in monitoring articulatory responses, see above), and sensorimotor cortices (necessary for implementing articulatory responses; see Introduction) (Dadario & Sughrue, 2023). From this perspective, our findings may reflect

goal-directed access to phonological information, perhaps to facilitate appropriate speech production.

We also detected phonological information in left occipital areas, including superior lateral occipital cortex, the occipital pole, and ventral occipital cortex. The latter (posteriormost) areas are sensitive to low-level orthographic properties (see first paragraph of this section); it is likely that such information is relevant to spelling-to-sound conversion. The role of superior lateral occipital cortex is not clear, though it is worth noting that immediately adjacent parietal cortices have been implicated in spelling-to-sound mapping (Levy et al., 2009).

Semantic information was decodable in right dorsal premotor cortex. This area has been implicated in a range of higher-order cognitive functions that may support integration of semantic information during articulation and / or episodic encoding. In particular, the cluster identified in our analysis approximately corresponds to the boundary of the dorsal and rostral (anterior) segments of dorsal premotor cortex, identified in an anatomical parcellation of this area by Genon et al. (2017). These authors suggest that the rostral segment in particular may be involved in higher-order processes, citing its functional connections to DLPFC and inferior parietal lobules. The role of DLPFC in language processing is discussed above, while the inferior parietal lobules have been linked to multisensory integration, particularly in the context of episodic encoding (Mesulam, 1998; Pasalar et al., 2010; Seghier, 2013; Zeller et al., 2015). As such, we posit that the role of dorsal premotor cortex in our study may reflect integration of each word's semantic features during articulatory planning or monitoring, via connections to DLPFC; recognizing that accessing semantic features of a word can facilitate access to that word's phonological representations (e.g., Bailey et al., 2023). Another possibility is that this area supports integration of semantic features into multisensory episodic representations (consistent with Fawcett et al., 2022), via connections to inferior parietal cortex.

3.7.2. Decoding Information In Silent Reading

We found that, relative to aloud reading, silent reading was associated with greater decodability of orthographic information in the right anterior temporal lobe. Notably, evidence for orthographic decodability increased in the posterior-to-anterior direction along the (anterior portion of the) ventral visual pathway (VVP)[11]. This finding is consistent with Zhao et al. (2017), who identified increasing decodability of orthographic information along the posterior-to-anterior axis of the VVP bilaterally. We interpret this finding in our study to reflect a shift towards direct print-to-meaning (i.e., orthographic-to-semantic) mapping, as opposed to phonologically-mediated access to meaning, during silent reading. While at face value this may seem like a contradictory statement—why would such an effect not be revealed by the semantic model?—we reason that that extraction of meaning directly from the visual word form should entail sensitivity to orthographic features in the area(s) responsible for that process (elaborated below). Importantly, our interpretation does not imply deeper semantic processing in the silent condition (as we did not observe silent > aloud semantic decodability); rather, it concerns the *routes* by which meaning is extracted.

Print-to-meaning mapping has been linked to connectivity along the VVP, particularly between the fusiform gyrus and anterior temporal regions (Taylor et al., 2017). The anterior temporal lobes in turn have long been associated with semantic access; in particular, the hub-and-spoke model of semantics (Patterson et al., 2007) considers these areas as a semantic 'hub' where inputs from multiple modalities (including visual word forms) converge. While recent versions of this model consider print-to-meaning operations to be left-lateralized (Ralph et al., 2017; Rice et al., 2015), some empirical work has supported right-sided contributions (Pobric et al., 2010; Taylor et al., 2017).

[11] The VVP runs from early visual cortex to anterior ventral temporal cortex, with subsequent projections to prefrontal areas (Kravitz et al., 2013).

Why should aloud and silent reading entail different degrees of print-to-meaning mapping? Connectionist models of reading consider access to meaning as the sum of parallel inputs from both direct and phonologically-mediated pathways (Harm & Seidenberg, 2004). From this perspective, the "division of labour" between pathways depends on ease of mapping in each—for example, sound-meaning mapping is ambiguous in the case of homophones (e.g., *ewes-use*), and may be resolved by assigning greater weighting to the direct pathway (Harm & Seidenberg, 2004). We reason that such division of labour may also be affected by task-related availability of phonological information. That is, if spelling-to-sound computations are already being promoted to meet task demands, as appears to be the case when reading aloud, then phonologically mediated access to semantics may be assigned greater weighting simply because the relevant computations are already being executed. By extension, reduced demands for phonological information when reading silently may result in *relatively* greater weighting on direct print-to-meaning mapping.

3.7.3. General Discussion

Our study provides evidence that aloud and silent reading entail differential decodability of multiple types of stimulus information. Broadly, this finding is consistent with the view of embodied and grounded cognition, whereby different experiences (e.g., reading aloud versus silently) give rise to fundamental changes in how stimuli are perceived and evaluated (Matheson & Barsalou, 2018). More specifically, we interpret our results to reflect flexible cognitive states during reading, whereby certain types of information are weighted according to the speakers' goals. Such goal-dependent weighting, we argue, is primarily driven by the demands of reading aloud.

How is such weighting determined? Consider that some types of information are more relevant or useful for certain tasks; consider also that any reading task will recruit cognitive resources that must be allocated in a manner which best serves the reader's goals. We suggest that, when reading aloud, phonological and articulatory information are weighted more heavily because those features are useful for planning, execution, and

monitoring of speech output. Similarly, visual information is necessary for recognising orthographic units; increased weighting here might therefore facilitate spelling-to-sound conversion. Semantic information may also facilitate access to phonological representations; alternatively, increased weighting of semantic information may reflect an intrinsic component of the encoding process.

When speech production is not required, the aforementioned types of information are less relevant to one's goal (reading the word), and so receive less weighting. We have argued that decreased weighting on phonological information in particular gives rise to greater weighting on orthographic information for the purpose of semantic access. More precisely, de-weighting of spelling-to-sound computations may necessitate relatively greater emphasis on direct print-to-meaning mapping when reading silently. Overall, our findings provide an adaptive view of information processing, whereby each type of information is weighted according to its task relevance.

A subtly different interpretation, and one which aligns with attentional accounts of the production effect (Fawcett, 2013; Fawcett et al., 2023; Mama et al., 2018; Ozubko, Gopie, et al., 2012), is overall reduced cognitive investment when reading words silently. Silent words might be processed relatively superficially—that is, readers can perceive and extract meaning from visually presented words based on orthographic information alone, without needing to invest further cognitive resources in phonological and articulatory computations. Such superficial processing may apply to all stimulus properties, rather than specific types of information that are directly relevant to production. Indeed, this provides a satisfying explanation for (relatively) reduced decodability of visual information in the silent condition: this information is arguably no more relevant to aloud reading than it is to silent reading, therefore decreased decodability may reflect a global decrease in cognitive investment.

Our findings also have implications for research on the production effect (and, more generally, for research on how words are encoded in memory). The production effect is reliably elicited when memory for aloud versus silent reading is tested, including

contexts in which participants are not informed in advance that they will be tested on the studied material (P. A. MacDonald & MacLeod, 1998; Zhou & MacLeod, 2021). Thus while we did not explicitly assess participants' memory for the words in the present experiment (largely because our stimulus list was relatively short, with each word repeated 4 times), the procedures employed in our experiment (other than stimulus repetition) accurately reflect the encoding conditions that reliably give rise to the production effect. Moreover, our findings align with theoretical accounts of the production effect. One account states that words read aloud are encoded as highly distinctive memory traces comprising sensorimotor information elicited during articulation (MacLeod et al., 2010). Our finding that reading aloud increased decodability of articulatory information appears to support this position. At the very least, we have shown that articulatory information is *available* for encoding at the neural level, which is a major assumption of distinctiveness accounts. As described in the Introduction, we do not feel that increased semantic decodability is incompatible with a distinctiveness-based account. It may be that the canonical memory advantage for words spoken aloud is facilitated both by distinctive sensorimotor experiences *and* enhanced semantic processing (a position previously expressed by Fawcett et al., 2022).

One possibility we must consider is that the results we observed are epiphenomenal to the production effect. Future research should therefore aim to link decodability (as measured in this study) directly to subjects' behavioural memory performance; this would provide a more complete understanding of which type(s) of information contribute to encoding. Moreover, future research might consider incorporating formal models of cognition. For example, the MINERVA-2 model of memory has proven able to reproduce the production effect by simulating sensory distinctiveness (Jamieson et al., 2016). In a different vein, some models consider cognitive representations as hierarchical, interconnected structures encompassing all types of stimulus information, as opposed to information being represented in terms of discrete processes (Matheson & Barsalou, 2018; Meyer & Damasio, 2009). Applying such models

in the context of RSA may provide a valuable means of testing formal cognitive theories of cognitive and neural representations.

3.8. Data And Code Availability

Code for all analyses reported in this manuscript is publicly available on GitHub [1]; additional materials that are necessary for analyses are stored in an Open Science Framework (OSF) repository [2]. Twelve participants consented to their anonymized data being made publicly available; raw data from those participants are available on the OSF repository [2]. Note that the data reported in this manuscript are from the "quickread" experiment described in both repositories.

[1] https://github.com/lbailey25/Production_Effect_MVPA

[2] https://osf.io/czb26/?view_only=86a66caf1d71484d8ef0293cfa2371df

3.9. Competing Interests Statement

The authors have no competing interests to declare.

3.10. Acknowledgements

We wish to thank the following individuals for their assistance with data collection: Matt Rogers, Laura McMillan, Cindy Hamon-Hill. We also wish to thank Philip Cook for his assistance with spatial normalisation of MVPA output. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to GEB and AJN (Grant numbers: RGPIN-2015-04131, RGPIN-2017-05340). LMB was supported by a Killam Predoctoral scholarship.

3.11. CRediT authorship statement

Lyam M. Bailey: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Project

administration. **Heath E. Matheson:** Conceptualization, Methodology, Software, Writing - review & editing. **Jonathan M. Fawcett:** Conceptualization, Writing - review & editing. **Glen E. Bodner:** Conceptualization, Funding acquisition, **Aaron J. Newman:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition.

3.13. Supplementary Materials

3.13.1 Dissimilarity measures used for construction of hypothesis models

For our visual measure, we first created an image of each word as it was presented in the experiment using the Pillow package (Umesh, 2012). Each image was then binarized (0 for background pixels, 1 for text pixels) and converted to a vector; we computed visual dissimilarity as the Pearson correlation distance between these vectors. For orthography, we took a similar approach to Fischer-Baum et al., (2017): features were unconstrained open character bigrams, computed using the wordkit package (Tulkens et al., 2018), and dissimilarity was calculated as Pearson correlation distance between feature vectors. For phonology, features were estimates of feedforward spelling-to-sound consistency obtained from Chee et al. (2020). Each word was represented as a vector of three values reflecting spelling-to-sound consistency estimates for the onset, nucleus, and coda of individual syllables. For multisyllabic words, we used Chee et al's (2020) composite scores, which reflect overall consistency estimates averaged across syllables. As such, we computed phonological dissimilarity as the Euclidean distance between vectors (we found that using Euclidean distances led to overall more normally distributed hypothesis models, compared with other metrics such as correlation or cosine distance).

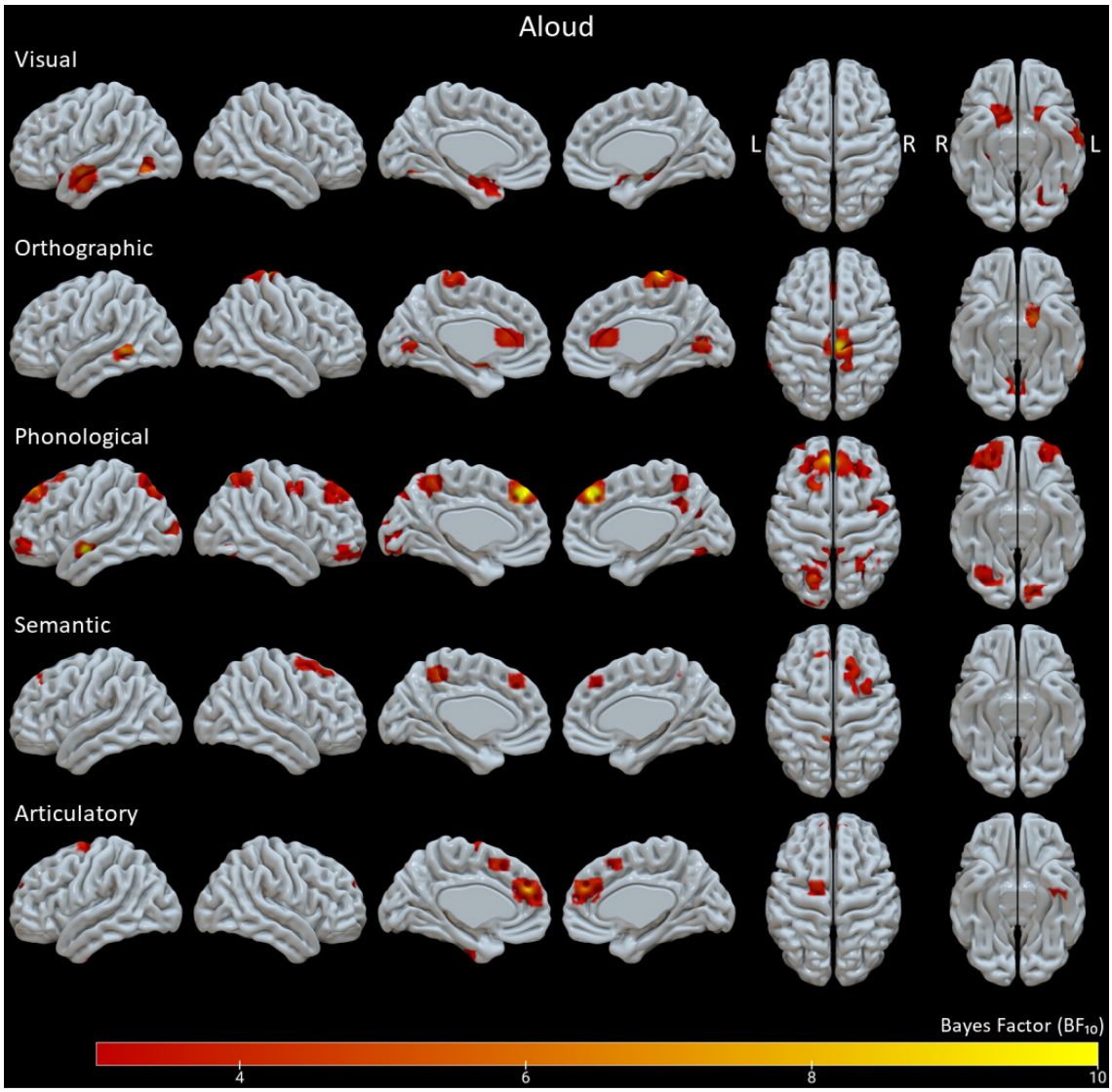
For semantic models, we used the gensim package (Řehůřek & Sojka, 2010) to extract vectors from the publicly available glove-wiki-gigaword-300 model (https://github.com/RaRe-Technologies/gensim-data); a pre-trained model of natural language processing based on the Global Vectors for Word Representation algorithm (GLoVe; Pennington et al., 2014). We computed semantic dissimilarity as the cosine distance between vectors. Articulatory models were created using the

PhonologicalCorpusTools package (Hall et al., 2019) in conjunction with the Irvine Phonotactic Online Dictionary (IPHOD; Vaden et al., 2009). Each word was transcribed into a string of its constituent phonemes; in turn, each phoneme was represented by a set of features as defined by Hayes (2008). Features reflected actions of the mouth, tongue, and vocal tract required to produce that phoneme (e.g., a phoneme may be voiced or voiceless, has a particular place of articulation, entails some degree of constriction by the lips, teeth, and vocal tract, etc.); thus, we considered them to reflect articulatory components of speech. Articulatory dissimilarity was computed as the edit distance between transcribed phoneme strings (that is, the number of additions, substitutions and/or deletions required to convert one phoneme string to another), weighted according to the number of features shared between constituent phonemes (Fontan et al., 2016). To control for confounding effects of word length, pairwise dissimilarity values were normalized by dividing the computed edit distance by the length of the longest word in the pair (as in Beijering et al., 2008; Schepens et al., 2012).

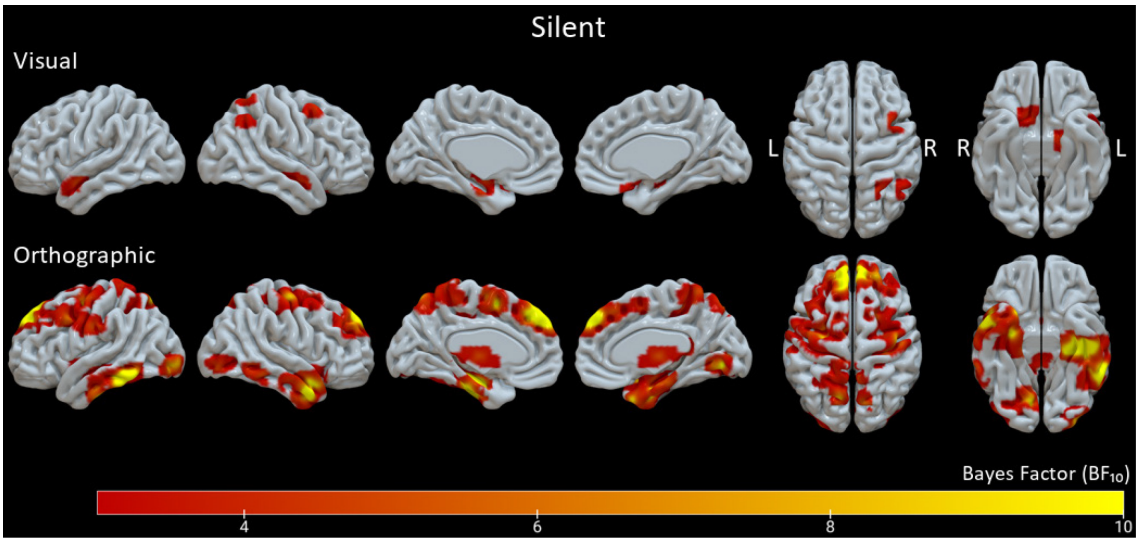
3.13.2. Supplementary Tables And Figures

Supplementary Table 3.1. Words used in the current study.

account	century	garden	language	pebble	summer
answer	department	handle	machine	powder	ticket
author	education	industry	message	record	turnip
beauty	envelope	journey	ocean	sailor	valley
campaign	forest	kingdom	painting	speech	wheat



Supplementary Figure 3.1. BF_{10} maps for the Aloud condition. Highlighted areas show evidence for greater-than-zero decodability. Surfaces from left to right show left lateral, right lateral, left medial, right medial, dorsal bilateral, and ventral bilateral views.



Supplementary Figure 3.2. BF_{10} maps for the Silent condition. Highlighted areas show evidence for greater-than-zero decodability. Surfaces from left to right show left lateral, right lateral, left medial, right medial, dorsal bilateral, and ventral bilateral views.

Supplementary Table 3.2. RSA results Table detailing clusters from the reading aloud condition. Clusters show evidence for greater-than-zero decodability.

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Visual	1	128	4.85	12.8	-36	-67	-4	**L Occipital Fusiform Gyrus (66)**
								L Inferior Temporal Gyrus temporooccipital (4)
								L Lateral Occipital Cortex inferior (58)
	2	307	4.28	9.69	-57	-6	-16	**L Middle Temporal Gyrus anterior (183)**
								L Insular Cortex (1)
								L Middle Temporal Gyrus posterior (17)
								L Planum Polare (6)
								L Superior Temporal Gyrus anterior (81)
								L Superior Temporal Gyrus posterior (15)
								L Temporal Pole (4)
	3	133	3.76	5.44	17	12	-18	**R Frontal Orbital Cortex (117)**
								R Parahippocampal Gyrus anterior (5)
								R Subcallosal Cortex (5)
								Right Putamen (13)
	4	86	3.57	5.56	-27	11	-25	**L Frontal Orbital Cortex (49)**
								L Insular Cortex (5)
								L Temporal Pole (32)
Orthographic	1	265	6.39	27.6	7	-29	74	**R Precentral Gyrus (151)**
								L Postcentral Gyrus (1)
								L Precentral Gyrus (25)
								R Postcentral Gyrus (88)
	2	42	6.36	13.1	-66	-46	-3	**L Middle Temporal Gyrus temporooccipital (31)**
								L Middle Temporal Gyrus posterior (11)
	3	30	4.09	8.6	13	-47	77	**R Postcentral Gyrus (20)**

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								R Superior Parietal Lobule (10)
	4	173	3.92	6.65	-4	27	11	**L Cingulate Gyrus anterior (122)**
								Left Lateral Ventricle (1)
								R Cingulate Gyrus anterior (16)
	5	66	3.89	5.26	3	-71	5	**R Lingual Gyrus (31)**
								L Lingual Gyrus (18)
								R Intracalcarine Cortex (17)
Phonological	1	122	9.19	42.5	-50	-8	-11	**L Superior Temporal Gyrus anterior (70)**
								L Middle Temporal Gyrus anterior (5)
								L Middle Temporal Gyrus posterior (1)
								L Planum Polare (36)
								L Superior Temporal Gyrus posterior (10)
	2	1080	6.6	39.5	0	40	46	**R Superior Frontal Gyrus (229)**
								L Frontal Pole (264)
								L Middle Frontal Gyrus (2)
								L Paracingulate Gyrus (2)
								L Superior Frontal Gyrus (310)
								R Frontal Pole (217)
								R Middle Frontal Gyrus (49)
								R Paracingulate Gyrus (5)
	3	222	5.63	19.3	-19	-90	2	**L Occipital Pole (106)**
								L Intracalcarine Cortex (23)
								L Lateral Occipital Cortex inferior (21)
								L Lateral Occipital Cortex superior (13)
								L Lingual Gyrus (10)
								L Occipital Fusiform Gyrus (49)
	4	212	5.01	15.8	21	-55	31	**R Precuneous Cortex (141)**

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								R Cingulate Gyrus posterior (29)
								R Cuneal Cortex (2)
								R Lateral Occipital Cortex superior (22)
								R Superior Parietal Lobule (2)
	5	160	4.23	6.99	36	-59	61	**R Lateral Occipital Cortex superior (135)**
								R Angular Gyrus (2)
								R Superior Parietal Lobule (23)
	6	23	4.21	6.84	-20	22	64	**L Superior Frontal Gyrus (23)**
	7	454	3.98	7.33	-16	-63	51	**L Lateral Occipital Cortex superior (289)**
								L Postcentral Gyrus (2)
								L Precuneous Cortex (114)
								L Superior Parietal Lobule (19)
								R Precuneous Cortex (30)
	8	140	3.72	5.62	27	48	-15	**R Frontal Pole (140)**
	9	61	3.57	5	45	-2	50	**R Precentral Gyrus (60)**
								R Middle Frontal Gyrus (1)
	10	118	3.42	4.41	-31	56	-7	**L Frontal Pole (118)**
	11	26	3.42	4.08	29	-72	-14	**R Occipital Fusiform Gyrus (26)**
Semantic	1	58	4.98	9.24	-10	-45	53	**L Precuneous Cortex (47)**
								L Postcentral Gyrus (11)
	2	188	3.9	6.96	27	16	63	**R Superior Frontal Gyrus (146)**
								R Middle Frontal Gyrus (42)
	3	51	3.83	5.5	-7	34	46	**L Superior Frontal Gyrus (49)**
								R Superior Frontal Gyrus (2)
Atriculatory	1	393	5.03	14	-1	46	25	**L Paracingulate Gyrus (212)**
								L Frontal Pole (5)
								L Superior Frontal Gyrus (27)

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								R Paracingulate Gyrus (121)
								R Superior Frontal Gyrus (28)
	2	23	3.91	6.89	-38	-14	-42	**L Temporal Fusiform Cortex posterior (19)**
								L Inferior Temporal Gyrus posterior (4)
	3	35	3.79	4.95	-18	-6	72	**L Superior Frontal Gyrus (35)**
	4	30	3.76	5.19	-3	17	51	**L Superior Frontal Gyrus (16)**
								L Paracingulate Gyrus (12)
								R Paracingulate Gyrus (1)
								R Superior Frontal Gyrus (1)

Supplementary Table 3.3. RSA results Table detailing clusters from the reading aloud condition. Clusters show evidence for greater-than-zero decodability.

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Visual	1	244	5.44	12.9	44	-6	-14	**R Planum Polare (108)**
								R Insular Cortex (49)
								R Middle Temporal Gyrus anterior (1)
								R Middle Temporal Gyrus posterior (6)
								R Superior Temporal Gyrus anterior (32)
								R Superior Temporal Gyrus posterior (41)
	2	112	4.03	7.05	-49	3	-19	**L Superior Temporal Gyrus anterior (27)**
								L Middle Temporal Gyrus anterior (9)
								L Planum Polare (20)
								L Temporal Pole (56)

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
	3	62	3.92	5.65	44	10	55	**R Middle Frontal Gyrus (62)**
	4	40	3.83	5.39	30	-56	68	**R Lateral Occipital Cortex superior (20)**
								R Superior Parietal Lobule (20)
	5	50	3.75	5.33	13	14	-22	**R Frontal Orbital Cortex (37)**
								R Subcallosal Cortex (13)
	6	30	3.59	4.83	50	-57	44	**R Lateral Occipital Cortex superior (15)**
								R Angular Gyrus (15)
Orthographic	1	2371	22.47	885	-50	-32	-22	**L Inferior Temporal Gyrus posterior (707)**
								L Inferior Temporal Gyrus anterior (10)
								L Inferior Temporal Gyrus temporooccipital (338)
								L Lateral Occipital Cortex inferior (3)
								L Middle Temporal Gyrus anterior (2)
								L Middle Temporal Gyrus posterior (262)
								L Middle Temporal Gyrus temporooccipital (67)
								L Parahippocampal Gyrus anterior (322)
								L Parahippocampal Gyrus posterior (125)
								L Planum Polare (11)
								L Superior Temporal Gyrus anterior (3)
								L Temporal Fusiform Cortex anterior (22)
								L Temporal Fusiform Cortex posterior (454)
								L Temporal Occipital Fusiform Cortex (20)
								Left Hippocampus (142)
	2	2454	14.71	251	-2	49	41	**L Superior Frontal Gyrus (607)**
								L Frontal Pole (635)
								L Paracingulate Gyrus (12)
								R Frontal Pole (898)
								R Middle Frontal Gyrus (9)

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								R Paracingulate Gyrus (9)
								R Superior Frontal Gyrus (280)
	3	2691	7.94	185	42	2	-30	**R Inferior Temporal Gyrus anterior (151)**
								Brain-Stem (1)
								R Frontal Orbital Cortex (4)
								R Inferior Temporal Gyrus posterior (23)
								R Insular Cortex (92)
								R Middle Temporal Gyrus anterior (284)
								R Middle Temporal Gyrus posterior (75)
								R Parahippocampal Gyrus anterior (429)
								R Parahippocampal Gyrus posterior (26)
								R Planum Polare (102)
								R Superior Temporal Gyrus anterior (69)
								R Superior Temporal Gyrus posterior (1)
								R Temporal Fusiform Cortex anterior (234)
								R Temporal Fusiform Cortex posterior (75)
								R Temporal Pole (965)
								Right Amygdala (32)
								Right Hippocampus (160)
	4	234	6.26	23.1	-34	-93	-6	**L Occipital Pole (154)**
								L Lateral Occipital Cortex inferior (79)
								L Occipital Fusiform Gyrus (1)
	5	584	5.73	36	22	-76	-4	**R Occipital Fusiform Gyrus (134)**
								R Intracalcarine Cortex (10)
								R Lateral Occipital Cortex inferior (170)
								R Lingual Gyrus (266)
								R Occipital Pole (4)

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
	6	2406	5.57	38.2	-2	-5	56	**L Supplementary Motor Area (246)**
								L Middle Frontal Gyrus (320)
								L Paracingulate Gyrus (2)
								L Postcentral Gyrus (12)
								L Precentral Gyrus (436)
								L Superior Frontal Gyrus (298)
								R Middle Frontal Gyrus (24)
								R Paracingulate Gyrus (7)
								R Postcentral Gyrus (233)
								R Precentral Gyrus (484)
								R Precuneous Cortex (1)
								R Superior Frontal Gyrus (268)
								R Superior Parietal Lobule (2)
								R Supplementary Motor Area (62)
	7	100	4.76	10.2	64	-48	-13	**R Middle Temporal Gyrus temporooccipital (47)**
								R Inferior Temporal Gyrus temporooccipital (52)
								R Middle Temporal Gyrus posterior (1)
	8	1321	4.32	17.5	-6	-57	62	**L Precuneous Cortex (242)**
								L Lateral Occipital Cortex superior (253)
								L Postcentral Gyrus (243)
								L Precentral Gyrus (39)
								L Superior Parietal Lobule (219)
								R Lateral Occipital Cortex superior (105)
								R Postcentral Gyrus (12)
								R Precuneous Cortex (208)
	9	43	4.21	6.86	12	-37	12	**R Cingulate Gyrus posterior (21)**
								R Precuneous Cortex (1)

Type of Information	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								Right Thalamus (3)
	10	185	4	6.97	-59	-11	35	**L Postcentral Gyrus (116)**
								L Precentral Gyrus (69)
	11	27	3.62	4.58	-38	0	-5	**L Insular Cortex (25)**
	12	23	3.36	4.14	-30	32	31	**L Middle Frontal Gyrus (23)**
	13	35	3.24	3.62	39	8	56	**R Middle Frontal Gyrus (35)**

3.14. Transition To Chapter 4 (Bridging Section)

Chapter 4 presents an under-review manuscript which, compared with Chapter 3, is more directly concerned with the production effect. This chapter concerns an experiment in which participants studied words by reading them aloud or silently and were later tested on their recognition memory for all the studied words. Therefore, similarly to Chapter 2, it was possible to link fMRI results directly to behavioural outcomes. Unlike Chapter 2, however, the experiment presented in Chapter 4 was designed to enable multivariate analysis of fMRI data. Chapter 4 investigates the presence of two phenomena—neural reactivation and transformation (discussed Section 1.7.)—in the context of the production effect. Much like Chapter 3, the group-level analysis presented in Chapter 4 depend on feature-wise Bayes factors (outlined in Section 2.13).

CHAPTER 4. DISSOCIABLE ROLES OF NEURAL PATTERN REACTIVATION AND TRANSFORMATION DURING RECOGNITION OF WORDS READ ALOUD AND SILENTLY: AN MVPA STUDY OF THE PRODUCTION EFFECT

4.1. Publication Information

Bailey, L.M., Matheson, H.E., Fawcett, J.M., Bodner, G.E., & Newman, A.J. (under review). Dissociable roles of neural pattern reactivation and transformation during recognition of words read aloud and silently: An MVPA study of the production effect. Manuscript ID: NSY-D-24-00076[12]

4.2. Abstract

Recent work surrounding the neural correlates of episodic memory retrieval has focussed on the decodability of neural activation patterns elicited by unique stimuli. Research in this area has revealed two distinct phenomena: (i) neural pattern reactivation, which describes the fidelity of activation patterns between encoding and retrieval; (ii) neural pattern transformation, which describes systematic changes to these patterns. This study used fMRI to investigate the roles of these two processes in the context of the production effect, which is a robust episodic memory advantage for words read aloud compared to words read silently. Twenty-five participants read words either aloud or silently, and later performed old-new recognition judgements on all previously seen words. We applied multivariate analysis to compare behaviourally relevant measures of reactivation and transformation between the two conditions. We found that, compared with silent words, successful recognition of aloud words was associated with reactivation in the left insula and transformation in the left precuneus. By contrast, recognising silent words (compared to aloud) was associated with relatively more extensive reactivation, predominantly in left ventral temporal and prefrontal areas. We suggest that recognition of aloud words might depend on retrieval and metacognitive evaluation of speech-related information that was elicited during the initial encoding experience, while recognition of silent words is more dependent on reinstatement of visual-orthographic information. Overall, our results demonstrate that different encoding conditions may give rise to dissociable neural mechanisms supporting single word recognition.

4.3. Statement Of Student Contributions To Manuscript

My contributions to this chapter included writing (original and subsequent drafts), study design, data collection, curation, analysis and interpretation, and creation of tables and figures.

[12] This chapter is a modified version of the manuscript submitted for peer review.

4.4. Introduction

Over the past two decades, developments in multivariate analyses of functional neuroimaging data have enabled researchers to decode information distinguishing individual stimuli from one another, based on the unique patterns of neural activation that they elicit (M. Peelen & Downing, 2022). This approach has been of great value to research examining neural correlates of episodic memory retrieval.

A major finding in this area is that the fidelity of activation patterns (within a particular brain area) to a unique study item between encoding and retrieval reliably predicts retrieval success for that item (e.g., Davis et al., 2014; Kuhl & Chun, 2014; Ritchey et al., 2013; Wing et al., 2015). This *reactivation* often occurs in high-level visual processing centres such as lateral occipital (LOC) and ventral temporal cortex (VTC) (Danker et al., 2017; Hasinski & Sederberg, 2016; Long & Kuhl, 2021; Ritchey et al., 2013; Wing et al., 2015; Xue et al., 2010; Zeithamova et al., 2017), as well as prefrontal and lateral parietal areas (Kuhl & Chun, 2014; Xiao et al., 2017; Xue et al., 2010, 2013; Zeithamova et al., 2017), and is generally taken to reflect visual object processing demands that were present during encoding, but absent (or only partially present) during retrieval[13]. For example, some studies have trained participants to learn word-picture pairings, and have shown that patterns elicited by the word-picture pair (or, alternatively, the picture alone) at encoding are reinstated during retrieval when participants are cued with the word and asked to recall its paired associate. Neural reactivation might therefore be attributed to faithful mental reinstatement of the original episode, grounded in the perceptual experiences that the episode entailed (e.g., Meyer & Damasio, 2009).

In contrast to reactivation, *neural transformation* describes systematic alteration of activation patterns elicited by the same stimulus. To our knowledge, transformation has been captured by prior literature in three main ways. First, stimulus-specific

[13] As an aside, we note that while visual object recognition is typically considered the domain of VTC and OTC, some multivariate studies have indicated that object category features may be decoded from patterns in frontoparietal cortices (e.g., Kuhl et al., 2013; Long & Kuhl, 2021; Zeithamova et al., 2017), indicating that these areas have the capacity to represent detailed stimulus information.

information that is decodable in perceptual cortices (e.g., visual areas) during encoding may be reinstated in frontoparietal regions during retrieval (Favila et al., 2018; Long & Kuhl, 2021; Xiao et al., 2017). Second, transformation has been captured as a change in activity patterns within a discrete region; one may measure the similarity of responses (to the same item) between two encoding episodes or two retrieval episodes, and also between encoding and retrieval. A decrease in similarity between encoding and retrieval, relative to encoding-encoding or retrieval-retrieval, may be taken as evidence of transformation within that region. Importantly, this second definition of transformation does not constitute a mere weakening of reactivation effects or an increase in noise, because item-specific information must still be decodable from the transformed (i.e., during retrieval) neural patterns (e.g., J. Chen et al., 2017; but see Xue, 2022 for review and discussion). This kind of transformation may therefore be conceptualised as a systematic change to the representational geometry (Kriegeskorte & Kievit, 2013) of stimulus-elicited activation patterns within a given region. Finally, transformation has also been captured as a change in the representational format of encoded information (i.e., the decodability of stimulus properties in activation patterns; Kriegeskorte et al., 2008). For example, some studies have characterised transformation as a shift from decodability of perceptual information to semantic information (Linde-Domingo et al., 2019; Liu et al., 2021). Broadly, transformation (by any of the above definitions) is thought to reflect reorganisation of encoded information, either as a natural part of the retrieval process or to meet specific retrieval goals (Favila et al., 2020; Xue, 2022). This perspective is consistent with the general view that remembering is more of a reconstructive than reproductive process (Schacter et al., 1998).

Reactivation and transformation may be relevant to the *production effect*, which is a robust recognition advantage for words read aloud compared to words read silently (MacLeod et al., 2010). Much work in this area emphasises the role of sensorimotor experiences conferred by verbalisation: movements of the tongue and jaw, vibration of the larynx, and auditory feedback from one's own voice. A popular view is that these distinctive experiences are appended to the episodic representation for having seen a

word, and subsequently may be leveraged to facilitate retrieval. For example, one account holds that, during retrieval, participants may recollect (i.e., mentally replay) the productive act; this enables them to reject or endorse the presented word (e.g., on an old/new judgement task) using a so-called distinctiveness heuristic: *I remember reading this word aloud, therefore I probably studied it in this experiment* (MacLeod & Bodner, 2017; Taikh & Bodner, 2016). In a similar vein, Wakeham-Lewis et al. (2022) recently proposed that when participants are presented with a word during a recognition task, they may mentally simulate reading that word aloud; the simulated experience may then be compared with the encoded episode to endorse or reject the presented word. Importantly, both of these accounts depend on access to encoded sensorimotor information that is elicited when reading aloud. At a neural level, both perspectives seem compatible with a reactivation-based perspective whereby retrieval entails recapitulation of perceptually-grounded experiences. That is, if distinctive sensorimotor experiences are mentally replayed or simulated (and then compared) at retrieval, retrieval ought to entail reactivation of neural patterns elicited during speech production. Remembering distinctive speech experiences may also entail neural transformation. Using a distinctiveness heuristic (MacLeod & Bodner, 2017), or comparing mentally simulated actions to encoded experiences (Wakeham-Lewis et al., 2022), may entail evaluation an/or reorganisation of encoded sensorimotor information—this would be consistent with work showing that the decodability of neural responses to the same stimulus can vary according to one's task (e.g., Favila et al., 2018; Nastase et al., 2017; Wang et al., 2018). Stated differently: speech-elicited information *for encoding* might manifest differently than *for retrieval*.

The goal of this study was to characterise how reactivation and transformation contribute to the production effect, using a whole-brain searchlight method[14]. Given the novelty of this work (to our knowledge, no prior work has investigated either process in

[14] This method parcellates a large area of cortex (in this case, the whole brain) into a series of smaller searchlight areas, and examinnes activation patterns within each area. As such, it allows one to effectively "scan" the entire brain, as opposed to examining a set of discrete ROIs.

the context of speech production), we were concerned that basing our analyses on a-priori regions of interest (ROIs) might blind us to potential areas of reactivation or transformation. As such, using a searchlight enabled fully data-driven identification of regions exhibiting interesting effects.

Participants read words aloud or silently (encoding phase), and later performed a recognition test (recognition phase), during fMRI scanning. We first computed behaviourally-relevant measures of reactivation and transformation using a whole-brain searchlight approach, for each individual subject. We quantified transformation as a systematic decrease in within-region pattern similarity (i.e., the second method of capturing transformation described earlier). Importantly, this measure was constrained by a between-subjects transformation analysis (J. Chen et al., 2017) which ensured that results were not driven by an increase in noise between encoding and recognition. Both measures were behaviourally relevant in that they quantified reactivation or transformation for subsequently remembered words relative to forgotten words; hence, they revealed areas in which either process was associated with successful recognition. We then compared, at the group level, reactivation and transformation between the aloud and silent reading conditions.

We predicted that, relative to words that were read silently, words that were read aloud would elicit reactivation in primary and associative sensorimotor and auditory cortices linked to speech production (Bailey et al., 2021; Dietz et al., 2005; Qu et al., 2022). Transformation effects may be present in these same speech-relevant areas, or alternatively in frontoparietal areas implicated by prior work on transformation between encoding and retrieval (J. Chen et al., 2017; Favila et al., 2018; Xiao et al., 2017). Although our hypotheses were mainly focussed on processes that were increased for aloud words (revealed by aloud > silent contrasts), it was also possible that reactivation or transformation might be increased during recognition of silent words (silent > aloud). Therefore, our analyses entailed bidirectional between-condition contrasts of reactivation and transformation.

4.5. Methods

4.5.1. Subjects

Thirty participants took part in this experiment, aged 18-40 (*M* = 21.43, *SD* = 4.57), 21 female, three left-handed. Participants were recruited through on-campus advertising at Dalhousie University, and received $30 CAD reimbursement and a digital image of their brain. Prior to taking part in the study, all potential participants were screened to ensure normal or corrected-to-normal vision, proficiency in English, no history of neurological illness or trauma, and no contraindications to MRI scanning. Handedness information was obtained using the Edinburgh Handedness Inventory (Oldfield, 1971). The experiment took place at the IWK Health Centre in Halifax, Canada, and all procedures were approved by the IWK Research Ethics Board. Participants provided informed consent according to the Declaration of Helsinki.

Of the 30 participants recruited, five were excluded from data analysis. One participant could not complete the experiment due technical problems; another disclosed that they did not fit inclusion criteria (neurological normality) after having completed the study. Two participants reported problems or errors completing the task as instructed: one reported that they silently mouthed words in the aloud condition instead of vocalising; the other had difficulty reading the presented words in the scanner. Finally, one participant made zero incorrect responses in at least one condition during the recognition phase, making them ineligible for our measures of reinstatement and transformation (which depended on contrasts between correct and incorrect responses). Therefore, data from 25 participants (17 female, 2 left-handed, *M* age = 21.40, *SD* = 4.73) are reported here.

4.5.2. Stimuli and Apparatus

Participants viewed all stimuli while lying supine in the MRI scanner; stimuli were presented using the VisualSystem *HD* stimulus presentation system (Nordic Neuro Lab, Bergen, Norway). During the experiment, participants made button-press responses using

MR-compatible ResponseGrip handles (Nordic Neuro Lab), one in each hand. Stimuli were presented using PsychoPy 2020.2.1 (Peirce et al., 2019); these appeared on an LCD screen positioned behind the scanner and viewed by participants via an angled mirror fixed to the MR head coil. Stimuli for the encoding and recognition tasks were 90 nouns selected from the list used in Bailey et al. (2021)[15], shown in Table 4.1. Words were 6 to 10 characters in length, each with a frequency greater than 30 per million (Thorndike & Lorge, 1944). Stimuli for the active baseline task (see below) were printed numbers 1-9. All stimuli (words and numbers) were presented at the centre of the screen, in white lowercase Arial font against a dark grey background (RGB: 128, 128, 128). During the encoding phase, participants were asked to respond to each presented word by either reading it aloud or reading it silently. Response cues were greyscale icons either depicting a mouth (read aloud) or an eye (read silently), presented at the centre of the screen.

[15] The only exception was the word *trousers* from Bailey et al. (2021); during piloting for the current study, multiple participants reported that this word "stood out" as unusual. Therefore, we replaced this word with *pants* in the current study.

Table 4.1. Stimulus words used in the current study

address	branch	distance	glass	justice	merchant	plate	shadow	treasure
afternoon	building	election	gravity	kettle	minute	pocket	shoulder	uncle
amount	capital	engine	guardian	kitchen	neighbour	porch	station	uniform
arrow	captain	entrance	harbour	knock	nephew	quarrel	steam	vacation
attention	castle	evening	history	ladder	office	quarter	stream	victory
attitude	clothes	factory	holiday	laugh	orchard	queen	teacher	village
avenue	daughter	fashion	invention	leather	package	resort	theatre	wagon
basket	debate	foundation	invitation	lesson	pants	reward	thread	wheel
battery	dinner	friend	island	market	partner	river	traffic	whisper
border	direction	furniture	judge	meadow	peace	school	travel	winter

4.5.3. Procedure

Participants first provided informed consent and completed MRI safety screening and the Edinburgh Handedness Inventory. Participants were informed that they would be taking part in a study on memory, and that they would be studying a list of words that they would later be tested on. Before entering the scanner, participants completed a shortened practice version of the main experiment, using a different stimulus set, on a laptop computer to familiarise them with the experimental tasks. After this practice task, a member of the research team confirmed (verbally) that participants understood the task instructions. Following this, participants were positioned in the MRI scanner. A brief scout scan determined head position. Participants then completed three functional scans followed by a structural scan. The first two functional scans (approximately 13 minutes each) comprised the *encoding phase*; the third (approximately 10 minutes) comprised the *recognition phase*.

At the start of each participant's scanning session, the 90 stimulus words (Table 4.1) were randomly allocated to one of three conditions: read aloud, read silently, or foil (30 words per condition). Words in the first two conditions were presented throughout the encoding phase, while all three conditions were present in the recognition phase. Stimulus-condition mappings were randomised between participants.

Encoding Phase. Each functional run of the encoding phase comprised 60 trials, and each word (from the read aloud and read silently conditions) was presented once per run (this meant that each word was seen twice during the study phase). Word presentation order was randomised for each run. Each trial began with a fixation cross ("+") presented for 500 ms, followed by a 1000 ms cue instructing participants to read the upcoming word either aloud or silently. The stimulus word for that trial was presented for 2500 ms, during which time participants were required to read the word as instructed. Following this, participants completed an active baseline task for 8000 ms. During the baseline task, a randomly generated number 1-9 appeared onscreen, and participants were instructed to indicate with a button press whether the number was odd or even.

Response mappings (left-hand button press for odd, right-hand for even) were presented in the top corners of the screen and were consistent across all participants. Each number was presented for 2000 ms, after which it was replaced with a different (randomly generated) number; this procedure was repeated until the end of the trial. Thus, the active baseline task for each trial comprised odd-even decisions on four numbers, each presented for 2000 ms. This procedure resulted in a trial SOA of 12 seconds. A schematic of encoding phase trial structure is shown in the left panel of Figure 4.1.

Recognition Phase. The recognition phase comprised a single run with 90 trials, in which all stimulus words were presented in random order. Sixty of these words were previously seen by participants in the encoding phase (30 aloud; 30 silent); the remaining 30 were unseen foils. Each trial began with a fixation cross for 1500 ms, followed by a stimulus word for 2500 ms. During this 2500 ms period, participants were instructed to indicate with a button press whether the presented word was “old” or “new”. Response mappings (left hand for old, right for new) were consistent across participants, and cues were presented in the top corners of the screen as reminders. The study word remained onscreen for the full 2500 ms regardless of when participants made a response. Following this, participants completed an identical active baseline task to that of the encoding trials before the start of the next trial. This procedure ensured a trial SOA of 12 seconds, as in the encoding phase. A schematic of recognition phase trial structure is shown in the right panel of Figure 4.1.

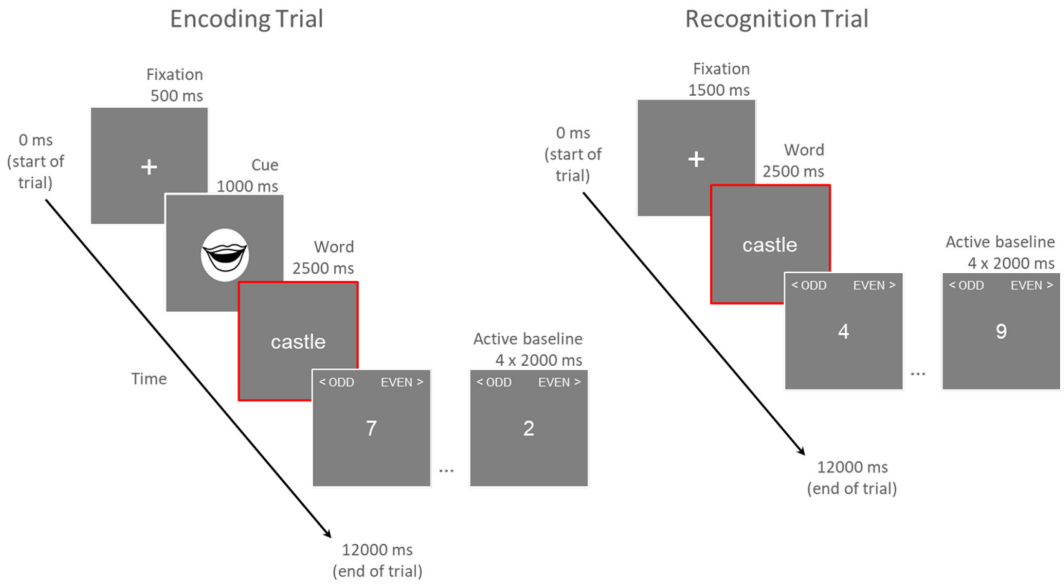


Figure 4.1. Schematic of an example trial from the encoding phase (left) and recognition phase (right). The red outline indicates the period modelled by the time series regressor for each trial (i.e., the temporal window from which activation patterns were estimated).

4.5.4. MRI Data Acquisition

MRI data was acquired on a 1.5 Tesla GE MRI system (GE Medical Systems, Waukesha, WI) equipped with a 19 channel head coil. Each participant completed three functional scans (described above) followed by a structural scan. Functional scans used a gradient echo-planar pulse sequence, TR = 1800 ms, TE = 23 ms, FOV = 240 mm, flip angle = 90°. Images were obtained in 34 axial slices[16] (no gap, sequential acquisition) of thickness = 3.75 mm, matrix = 64 x 64, resulting in an in-plane voxel resolution of 3.75 x 3.75 mm. The FOV included full cortical coverage and partial cerebellar coverage. We collected 400 functional volumes for each encoding run; 600 for the recognition run. Five additional volumes were acquired at the start of each run, but were discarded following acquisition. For the structural scan, a T1-weighted anatomical image was obtained using

[16] We added 4 axial slices (total = 38) to the protocol for one participant, in order to accommodate their entire cerebral cortex.

a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence, TI = 1134 ms, flip angle = 8°, NEX = 2, FOV = 224 mm, matrix = 224 x 224, resulting in an in-plane voxel resolution of 1 x 1 mm.

4.5.5. fMRI Data Processing

fMRI data were preprocessed using functions from FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). FSL functions were implemented with custom bash shell scripting, in parallel for multiple subjects/runs using GNU *parallel* (Tange, 2011) to improve efficiency. Preprocessing steps included non-brain removal with BET (S. M. Smith, 2002), grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor, high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 50.0 s), spatial smoothing (FWHM Gaussian kernel, 6 mm), and motion correction using MCFLIRT (Jenkinson et al., 2002). Our a priori threshold for excessive head motion between contiguous time points was 2 mm; no participant exceeded this threshold on any run.

We extracted trial-level activation patterns using the least-squares-all (LSA) method (Mumford et al., 2014). Using FEAT, we modelled each of the three functional runs using a GLM with one regressor per trial. Each regressor comprised a time series modelling the period during which the word for that trial was presented (duration: 2500 ms), convolved with a gamma function (lag: 6 sec, sigma: 3 s) as a model of the haemodynamic response function (HRF). Each GLM also included parameters from motion correction as regressors of no interest. For each model, we defined contrasts of interest whereby each contrast comprised one regressor (i.e., trial) in the model. As a result, each participant was associated with 210 whole-brain functional maps (contrasts of parameter estimates; COPEs) in native functional space, with each map corresponding to a unique trial (60 in each encoding run; 90 in the recognition run). We next spatially transformed these functional maps to the MNI152 template. First, each participant's example functional volume (that is, a functional image acquired in the middle of each run)

was rigidly aligned to their high-resolution structural T1 image. Example functional volumes were taken from whichever run the to-be-transformed functional map belonged to. Each structural image was also aligned to MNI152 space with an affine transform. We then combined these EPI-to-structural and structural-to-MNI152 transforms to generate a matrix for transforming data from native functional space to MNI152, for each subject and run. We applied these matrices to the functional maps derived from the GLMs, such that each participant was associated with 210 functional maps in MNI152 space, re-sliced to 2 mm isotropic resolution. These transformation steps were performed using FSL's FLIRT function (Jenkinson et al., 2002; Jenkinson & Smith, 2001). For convenience, these preprocessing stages were applied to words in all conditions, however we discarded output for foil words (from the recognition phase) from subsequent fMRI analyses.

4.5.6. Subject-Level Analyses

We computed behaviourally-relevant measures of reactivation and transformation (described in the following sections) and implemented each using a whole-brain searchlight (spherical searchlight area, radius = 3 voxels, average volume = 115.4 voxels). These measures were behaviourally-relevant in that they quantified the difference in reactivation or transformation between subsequently remembered and forgotten items. Both measures considered correlations between activity patterns in the same local region (i.e., same searchlight sphere) across encoding and recognition. Searchlights for each measure were implemented independently for each participant and condition (aloud or silent). The searchlights and all other procedures described in this section were carried out in the MATLAB environment using functions from the CoSMoMVPA toolbox (Oosterhof et al., 2016) and custom scripting.

Reactivation. Reactivation analyses were based on the approach used by Zeithamova et al. (2017). These authors devised a measure for what they described as a *subsequent memory effect*; this measure was intended to capture larger same-item pattern similarity for subsequently remembered stimuli relative to subsequently

forgotten stimuli, as observed in previous literature (see Introduction). For convenience, we will refer to this measure in our study as a *reactivation index*, recognising that it still captures the relationship between pattern similarity and subsequent memory. First, we sorted activation patterns (contrasts of parameter estimates from the GLMs described in Section 4.5.5) according to the run (Encoding 1, Encoding 2, or Recognition) and stimulus to which they corresponded. Next, in each searchlight sphere, we computed pairwise correlations between patterns elicited by the same item across both encoding runs and the recognition run. That is, for each item, we computed *r*(Encoding 1, Recognition) and *r*(Encoding 2, Recognition); correlation values derived from both encoding runs were included in the reactivation index calculation (described below). All correlation values were sorted according to whether the eliciting stimulus was subsequently remembered (correctly responded "old" during the recognition task) or forgotten (incorrectly responded "new") by that participant. Following this, we computed a reactivation index as the standardised mean difference (Cohen's D) between remembered same-item correlations and forgotten same-item correlations. This calculation is illustrated in Figure 4.2. This procedure yielded a whole-brain map of reactivation indices, and was implemented independently for each participant and condition (aloud or silent).

Transformation. We performed the transformation analysis in a manner that was as methodologically consistent as possible with the reactivation analysis. We broadly conceptualised transformation as the difference between encoding-recognition similarity and encoding-encoding (same item, between runs) similarity within each searchlight sphere. As such, we computed a *transformation index* as the difference (of encoding-recognition / encoding-encoding differences) between subsequently remembered and forgotten items—therefore, our transformation index was behaviourally relevant in the same way as our reactivation index.

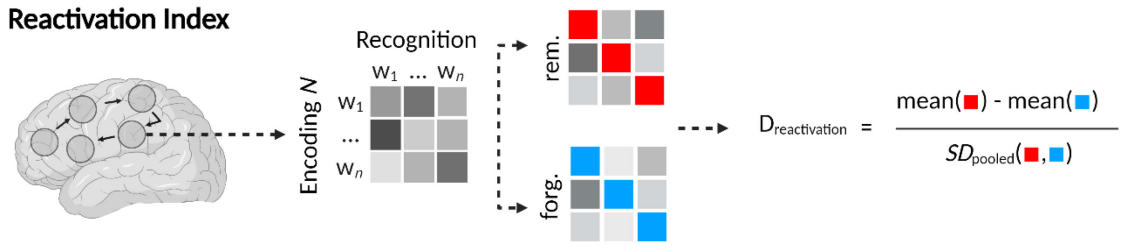


Figure 4.2. A reactivation index was computed at each searchlight centre. We first calculated pairwise correlations for each word ($w_{1 \ldots n}$) between each encoding run (Encoding N) and the recognition run (Recognition). Cells (squares) in the grayscale matrices represent correlations between pairs of words; on-diagonal cells contain same-item correlations used in this analysis (off-diagonal cells are not relevant to this analysis, and were therefore not computed, but are shown in the illustration to facilitate conceptual understanding). Same-item correlations (on-diagonal cells) were sorted according to subsequent memory in the recognition phase: either subsequently remembered (rem.) or subsequently forgotten (forg.). A reactivation index ($D_{reactivation}$) was computed as the standardised mean difference between correlations for correct items (red squares) versus correlations for incorrect items (blue squares).

Within each searchlight sphere, we first computed same-item correlations for encoding-encoding and encoding-recognition pairings, separately for subsequently remembered and forgotten items (as with the reactivation analysis, we considered correlations for both encoding runs relative to recognition). Following this, for each item, we subtracted each encoding-recognition correlation value from the encoding-encoding correlation value. The reasoning behind this step is that if patterns are transformed between encoding and recognition, encoding-recognition correlations should be smaller than encoding-encoding correlations, meaning that the product of the above subtraction should be a positive value. Moreover, the magnitude of this value should increase with larger differences between encoding-encoding and encoding-recognition (i.e., larger values correspond to greater degrees of transformation). This step generated two sets of values (one set per response type: subsequently remembered or forgotten), wherein every value reflected the difference in correlations between encoding-encoding and encoding-recognition for a single unique item. We computed a transformation index as

the normalised mean difference (Cohen's D) in correlation-difference values between subsequently remembered and subsequently forgotten items. This procedure is illustrated in Figure 4.3 (upper panel). As before, we generated maps of transformation indices independently for each participant and condition. Hereon in, we will refer to this measure as the *within-subjects transformation index*.

To verify that our within-subjects index captured true transformation—that is, systematic changes in activation patterns, rather than mere weakening of encoding patterns or an increase in noise (Xue, 2022)—we constrained the results of our within-subjects analysis with those of a between-subjects analysis (similar to J. Chen et al., 2017). To be clear: results from the between-subjects analysis are not reported here: rather, they were simply used to mask results of the within-subjects analysis. The between-subjects analysis allowed us to identify areas in which recognition-recognition correlations (across participants) were stronger than encoding-recognition correlations (also across participants). If apparent transformation effects were driven by weakened / more noisy response patterns, stimuli should not be more decodable at recognition (Xue, 2022). Stated differently: if systematic transformation had occurred, patterns elicited (in different participants) by the same item during the recognition phase should be more highly correlated than patterns elicited between encoding and recognition.

We computed a *between-subjects transformation index* at each searchlight centre, for each participant. This index was based on correlations of activity patterns between pairs of participants (see below), therefore our calculations were necessarily constrained to items that were shared across pairs (which varied because words were randomly assigned to aloud or silent reading). For this reason, we only considered items that were subsequently remembered by both subjects in each pair; it was not feasible to compute a remembered > forgotten between-subjects index, as very few forgotten items were shared across participants.

First, for each participant, we iteratively computed same-item correlations between recognition patterns from that (*k*th) participant and patterns at encoding[17] or recognition from each *other* (*j*th) participant. This yielded between-participant values for $r(\text{recognition}_k, \text{encoding}_j)$ and $r(\text{recognition}_k, \text{recognition}_j)$. We then computed (for each participant) a between-subjects transformation index as Cohen's D for the difference between $r(\text{recognition}_k, \text{recognition}_j)$ values and $r(\text{recognition}_k, \text{encoding}_j)$ values. This calculation is illustrated in Figure 4.3 (lower panel).

4.5.7. Group-Level Behavioural Analyses

Behavioural analyses were performed in the R environment (version 1.0.4; R Core Team, 2021). We analysed participants' button-press responses from the recognition task. During this task, participants saw words that they had either read aloud or silently during the encoding phase, or had not previously seen (foil words). On each trial, participants indicated whether the presented word was "old" or "new". Performance was measured as the percentage of recognition trials on which participants made an "old" response; statistical analyses were performed on mean percentages of "old" responses in each condition from each participant. Statistical analysis between conditions was based on Bayes Factors (BFs), computed in the R environment using the BayesFactor package (Morey et al., 2022). A Bayes Factor BF_{10} is a ratio expressing the likelihood of the data under an alternative hypothesis (H_1) relative to the null hypothesis (H_0). Within this framework, increasing BF_{10} values > 1.0 correspond to increasing strength of evidence for H_1 (Dienes, 2014; M. D. Lee & Wagenmakers, 2014). We tested the hypothesis H_1 that the aloud condition elicited more "old" responses compared to the silent and foil conditions, with separate paired-samples two-tailed Bayes *t*-tests, using default JZS priors (Rouder et al., 2009); this yielded a single Bayes Factor for each between-condition contrast.

[17] To reduce computational load, for each participant we first averaged together activity patterns from the two encoding runs for each item. As such, this analysis did not include separate correlations for each encoding run, as did the reactivation and within-subjects transformation analyses.

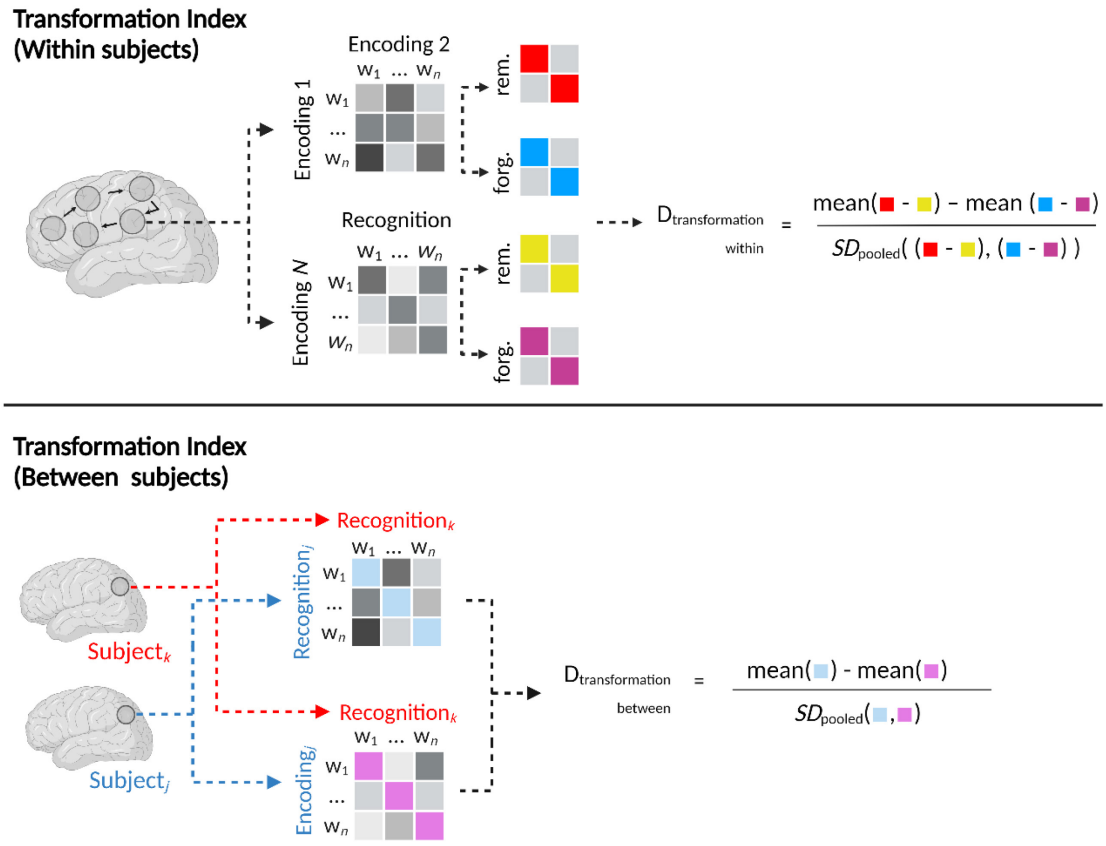


Figure 4.3. Within- and between-subjects transformation indices were calculated at each searchlight centre, for each individual subject. **Within subjects:** we calculated pairwise correlations for each word $w_{1 \ldots n}$ between the two encoding runs [*r*(Encoding 1, Encoding 2)], separately for subsequently remembered items (rem., red squares) and forgotten items (forg., blue squares). We also calculated pairwise correlations between each encoding run and the recognition run [*r*(Encoding N, Recognition)], again for remembered (gold squares) and forgotten (purple squares) items. Correlation difference values were computed as [*r*(Encoding 1, Encoding 2) - *r*(Encoding N, Recognition)], separately for remembered and forgotten items. A within-subjects transformation index ($D_{transformation\ within}$) was then computed as the standardised mean difference between remembered correlation difference values and forgotten correlation difference values. **Between subjects:** activity patterns from each subject *k* were iteratively compared to every other subject *j*. This analysis only considered correctly remembered items that were shared by *k* and *j*. We computed same-item correlations for each word $w_{1 \ldots n}$ between $encoding_k$ (patterns were averaged across encoding runs) and $retrieval_j$, and between $retrieval_k$ and $retrieval_j$. A between-subjects transformation index ($D_{transformation\ between}$) was computed as the standardised mean difference between r($recognition_k$, $recognition_j$) and r($encoding_k$–$recognition_j$).

4.5.8. Group-Level Multivariate fMRI Analyses

We performed group-level analyses on participants' searchlight output based on feature (voxel)-wise Bayes Factors, in line with the application of Bayes Factors in previous multivariate neuroimaging studies (Grootswagers, Robinson, & Carlson, 2019; Grootswagers, Robinson, Shatek, et al., 2019; Kaiser et al., 2018; Matheson et al., 2023; Moerel et al., 2022; Proklova et al., 2019; Teichmann et al., 2022). We performed voxel-wise Bayes *t*-tests in the MATLAB environment with functions from the bayesFactor package (Krekelberg, 2022), using default JZS priors (Rouder et al., 2009). We first tested for the presence of reactivation and transformation (both within- and between-subjects) in each condition alone. At each searchlight centre we performed a one-sample right-tailed Bayes *t*-test for the hypothesis H_1 that reactivation/transformation indices were greater than zero. This procedure was repeated for each condition and measure, yielding six whole-brain statistical maps of BF_{10} values. Results from these within-condition analyses are not explicitly reported here (instead they are presented in Supplementary Figures 4.1–4.3, and Supplementary Tables 4.1 and 4.2); rather, they were used to constrain results from the between-condition analyses (described as follows). Next, we sought to test the hypothesis H_1 that there was a difference in either reactivation or within-subject transformation between the aloud and silent conditions. At each searchlight centre, we performed a two-tailed Bayes *t*-test for a difference between conditions. This procedure was repeated independently for each measure, thus generating two statistical maps of BF_{10} values.

We thresholded all group-level statistical maps at $BF_{10} \geq 3.00$, thereby constraining results to voxels expressing at least moderate evidence for H_1 (consistent with conventional qualitative benchmarking of Bayes factors; M. D. Lee & Wagenmakers, 2014). In the context of a whole-brain analysis, we do not feel that reporting weak evidence ($BF_{10} < 3$) would be particularly valuable, as it would distract from areas where there was relatively stronger evidence for H_1. To determine directionality of between-condition effects concerning reactivation and within-subjects transformation, we calculated two average contrast maps for each of those measures. An aloud > silent map

was computed by subtracting the average searchlight results (across all participants) in the silent condition from those of the aloud condition. A silent > aloud map was computed in the opposite direction. These average contrast maps were used to mask the thresholded BF_{10} maps, meaning that each measure was now associated with two BF_{10} maps: one reflecting voxels where group-average values were higher in the aloud condition compared to silent (aloud > silent), and one vice-versa (silent > aloud). We further constrained the between-condition BF_{10} maps by masking each with the one-sample *t*-test map for its respective minuend condition (e.g., the aloud > silent reactivation map was masked with the one-sample reactivation map for the aloud condition). This ensured that areas showing differences between conditions were not driven entirely by negative values in the subtrahend condition.

Finally, we masked each within-subjects transformation contrast (aloud > silent or silent > aloud) with the between-subjects map for its corresponding minuend condition. Our choice to perform masking with individual conditions (as opposed to contrasts between conditions) is based on the following reasoning. The between-subjects measure was intended to capture transformation as conceptualised in previous literature (i.e., not driven by noise or weakened responses) (Xue, 2022). It was *not* intended to capture behaviourally-relevant differences that might exist between aloud and silent reading. As such, this analysis simply served to verify whether areas implicated in the (behaviourally relevant) between-condition contrasts also exhibited "true" transformation in their respective minuend condition.

We generated interpretable tables of results for each contrast (as well as individual conditions; see Supplementary Materials) using FSL's *cluster* and *atlasquery* functions, and custom bash scripting. Tables report spatial extent, mean and maximum BF_{10} values, and centre of gravity (COG) MNI coordinates for clusters of contiguous voxels (minimum spatial extent = 20 voxels). Mean and maximum BF_{10} are arguably both informative—local maxima are typically reported in fMRI studies, and in this context reflect the strongest available evidence for H_1 in a given cluster. On the other hand, mean BF_{10} values are more representative of evidential strength when considering whole

clusters. Anatomical labels for each cluster were identified from the Harvard-Oxford cortical and subcortical atlases.

4.6. Results

4.6.1. Behavioural Responses

Condition-wise group means for the percentage of "old" responses on the recognition task are illustrated in Figure 4.4. The results of independent Bayes *t*-tests for differences between conditions revealed strong evidence for a higher proportion of "old" responses in the aloud condition relative to the silent (BF_{10} = 11.33) and foil (BF_{10} > 10,000) conditions.

4.6.2. Encoding-Recognition Reactivation

Results for the reactivation analysis are presented in Table 4.2 and Figure 4.5. The aloud > silent contrast revealed a single cluster with strong evidence (Mean, Maximum BF_{10} = 11.04, 38.40) situated in the posterior portion of the left insula.

The silent > aloud contrast revealed more extensive reactivation, with clusters throughout frontal, parietal, and ventral temperooccipital cortices. In the frontal lobe, clusters with strong evidence (Mean $BF_{10} \geq 10$) for this contrast were present in the superior and middle frontal gyri bilaterally, as well as the left pars opercularis and supplementary motor area. Frontal clusters with relatively weaker evidence (that is, Mean BFs closer to or < 10) were present in the frontal poles bilaterally. In the parietal lobe, clusters with strong evidence were present in the postcentral gyrus bilaterally; one additional parietal cluster with moderate evidence was situated in the right precuneus. In ventral temperooccipital cortex, clusters with strong evidence were present bilaterally, in the temperooccipital and (more posterior) occipital portions of the fusiform gyrus. Notably, the cluster with the strongest evidence for this contrast was situated in left occipital fusiform cortex (Mean, Maximum BF_{10} = 86.47, 1430).

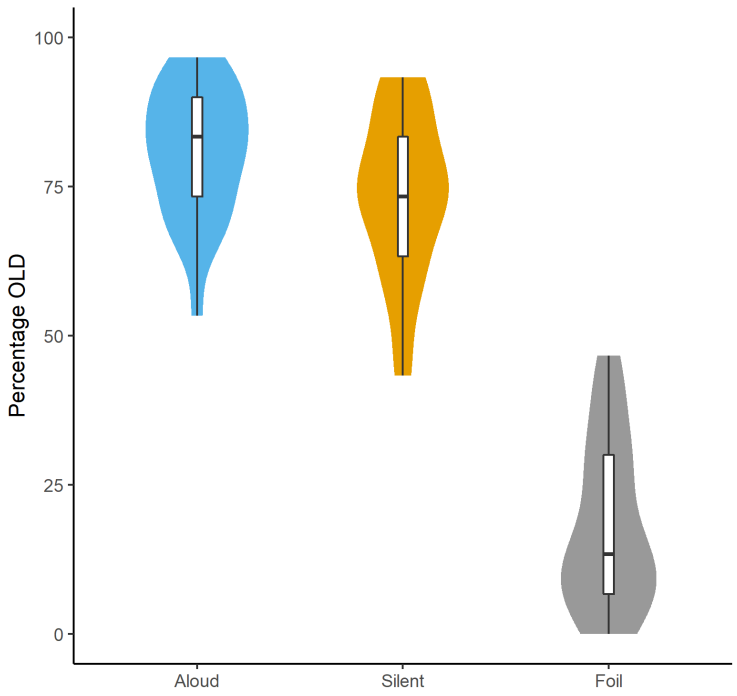


Figure 4.4. Percentages of OLD responses to words in each condition during the recognition task. Violin plots show the distributions of individual participant percentages in each condition. Box plots show condition-wise means across participants (middle bar), upper and lower quartiles (upper and lower box limits respectively) and ranges (whiskers).

Table 4.2. Clusters exhibiting evidence of between-condition differences in reactivation. Anatomical labels in boldface reflect the structure in which the centre of mass (COM) for that cluster was located. Values in parentheses for each anatomical label indicate the number of searchlight centres (i.e., voxels) within that label contributing to the cluster.

Contrast	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Aloud > Silent	1	49	11.04	38.4	-34	-20	15	**L Insular Cortex (37)**
								L Central Opercular Cortex (5)
								Left Putamen (1)
Silent > Aloud	1	108	86.47	1430	-23	-80	-8	**L Occipital Fusiform Gyrus (82)**
								L Lateral Occipital Cortex inferior (14)
								L Lingual Gyrus (12)
	2	112	73.58	991	-17	4	71	**L Superior Frontal Gyrus (95)**
								L Middle Frontal Gyrus (17)
	3	38	58.1	296	-42	11	19	**L Inferior Frontal Gyrus pars opercularis (36)**
								L Precentral Gyrus (2)
	4	62	51.81	588	10	36	46	**R Superior Frontal Gyrus (46)**
								R Frontal Pole (16)
	5	22	37.19	129	-6	-40	75	**L Postcentral Gyrus (22)**
	6	35	32.4	291	39	-55	-19	**R Temporal Occipital Fusiform Cortex (33)**
								R Inferior Temporal Gyrus temporooccipital (2)
	7	25	28.92	215	17	-33	46	**R Postcentral Gyrus (10)**
								R Precentral Gyrus (14)
								R Precuneous Cortex (1)
	8	70	24.81	177	-15	-9	46	**L Supplementary Motor Area (39)**
								L Cingulate Gyrus anterior (1)
								L Cingulate Gyrus posterior (1)
								L Precentral Gyrus (3)
								L Superior Frontal Gyrus (5)

Contrast	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
	9	195	21.13	285	-42	-60	-15	**L Temporal Occipital Fusiform Cortex (62)**
								L Inferior Temporal Gyrus temporooccipital (40)
								L Lateral Occipital Cortex inferior (43)
								L Middle Temporal Gyrus temporooccipital (9)
								L Occipital Fusiform Gyrus (41)
	10	179	17.87	340	-17	30	46	**L Superior Frontal Gyrus (135)**
								L Frontal Pole (11)
								L Middle Frontal Gyrus (33)
	11	52	12.15	105	-16	60	-18	**L Frontal Pole (52)**
	12	25	10.77	32.2	-47	37	24	**L Frontal Pole (13)**
								L Middle Frontal Gyrus (12)
	13	46	8.82	38.3	39	-84	-12	**R Lateral Occipital Cortex inferior (44)**
								R Occipital Pole (2)
	14	28	7.59	18.9	19	51	-7	**R Frontal Pole (26)**
								R Frontal Medial Cortex (1)
	15	24	6.87	14	7	60	22	**R Frontal Pole (24)**
	16	44	5.97	12.7	5	-59	44	**R Precuneous Cortex (44)**
	17	53	5.86	27.7	-25	49	24	**L Frontal Pole (53)**

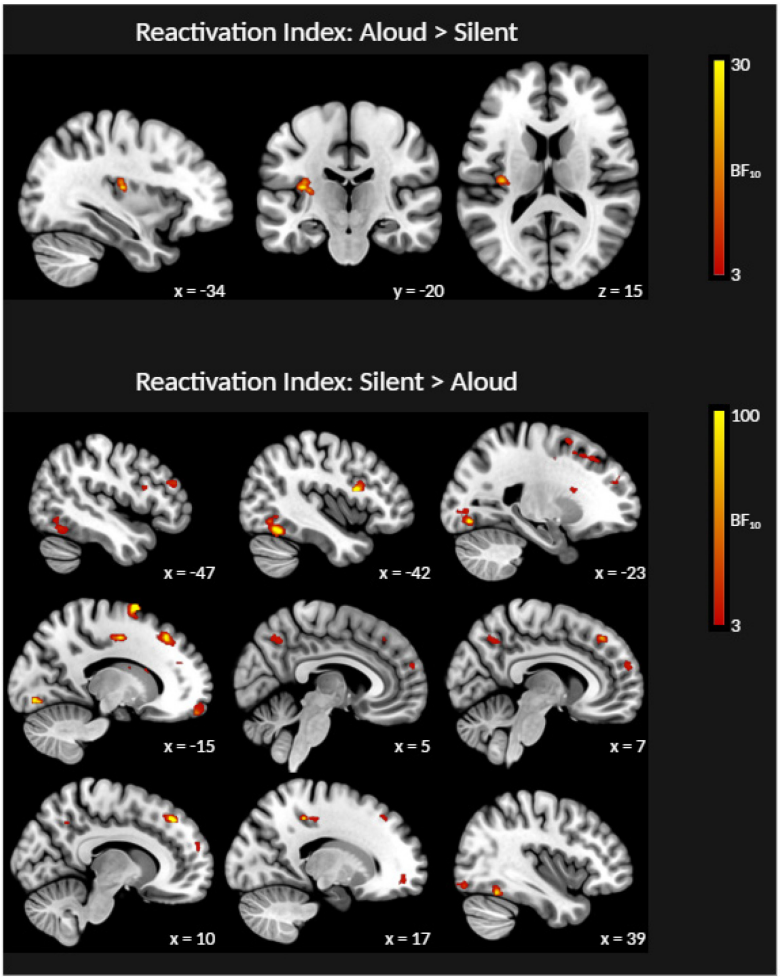


Figure 4.5. Cross-sectional slices show BF_{10} maps from the reactivation analyses. Highlighted areas show evidence of greater reactivation in the aloud reading condition relative to the silent reading condition (top panel) and vice-versa (bottom panel). Axial images are in neurological orientation (L-R).

4.6.3. Encoding-Recognition Transformation

Results for the transformation analysis are presented in Table 4.3 and Figure 4.6. The aloud > silent contrast revealed one cluster with strong evidence (Mean, Maximum BF_{10} = 991.56, 28000), situated in the posterior portion of the left precuneus. Another cluster with relatively weaker evidence (Mean, Maximum BF_{10} = 6.46, 11.70) was situated in the posterior portion of left ventral temporal cortex. The silent > aloud contrast did not reveal any clusters.

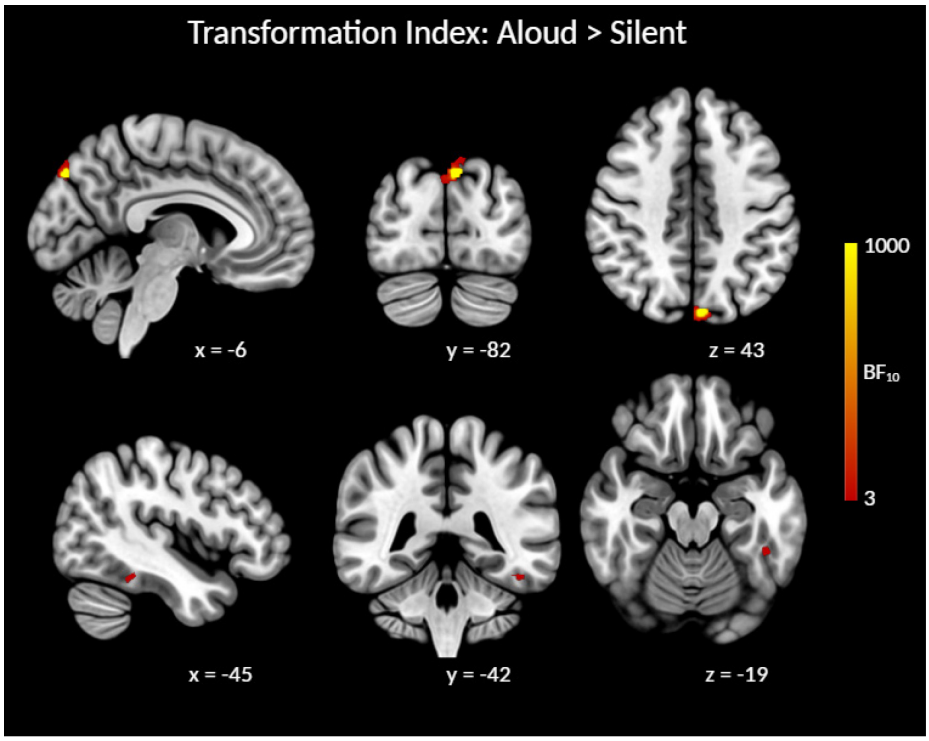


Figure 4.6. Cross-sectional slices show the BF_{10} map from the aloud > silent contrast of within-subjects transformation indices, masked with the map from the aloud > 0 contrast of between-subjects transformation indices. Highlighted areas show evidence of greater transformation in the aloud reading condition relative to the silent reading condition. Axial images are in neurological orientation (L-R).

Table 4.3. Clusters exhibiting evidence of between-condition differences in transformation. Anatomical labels in boldface reflect the structure in which the centre of mass (COM) for that cluster was located. Values in parentheses for each anatomical label indicate the number of searchlight centres (i.e., voxels) within that label contributing to the cluster.

Contrast	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Aloud > Silent	1	59	991.56	28000	-6	-82	43	**L Precuneous Cortex (17)**
								L Cuneal Cortex (7)
								L Lateral Occipital Cortex superior (33)
								R Cuneal Cortex (2)
	2	26	6.46	11.7	-45	-42	-19	**L Inferior Temporal Gyrus posterior (13)**
								L Inferior Temporal Gyrus temporooccipital (1)
								L Temporal Fusiform Cortex posterior (11)
								L Temporal Occipital Fusiform Cortex (1)

4.7. Discussion

This study builds upon prior work exploring neural correlates of the production effect (Bailey et al., 2021; Hassall et al., 2016; B. Zhang et al., 2023), and is the first such study to offer insights about its underlying spatial dynamics from a multivariate perspective. Our behavioural analyses showed that we replicated the standard production effect: participants remembered a larger proportion of words from the aloud condition relative to the silent condition. With respect to fMRI analyses, we investigated relative contributions of neural pattern reactivation and transformation during recognition of words from the aloud and silent reading conditions. Our measures of reactivation and transformation quantified differences between subsequently remembered and forgotten items, therefore they inherently capture processes underlying recognition success. We compared reactivation and transformation between the two reading conditions, thus revealing areas in which these processes were differentially involved in remembering.

We predicted that we would observe relatively more reactivation and/or transformation for aloud words relative to silent words. While this was true for transformation—which was exclusively detected by the aloud > silent contrast—we in fact observed relatively more widespread reactivation in the silent > aloud contrast (evidenced by a larger number or clusters with generally stronger evidence) compared with the aloud > silent contrast. These results indicate that remembering words that were read aloud versus silently seems to depend, to some extent, on dissociable neural mechanisms. This is in contrast to current theoretical accounts of the production effect which are (necessarily) one sided; these accounts generally frame encoding and subsequent retrieval as being stronger or more elaborate for aloud words (e.g., Fawcett et al., 2022; Jamieson et al., 2016; MacLeod & Bodner, 2017), with little consideration of processes that might be enhanced for silent words. Stated differently, aloud reading is generally considered to entail all the same processes as silent reading, *plus* processes elicited by speech production. Our results provide a rather different perspective: that word recognition may entail different mechanisms, depending on the preceding reading

conditions, rather than greater or lesser degrees of the same mechanism(s). Below we elaborate on the functional significance of reactivation and/or transformation effects identified by each contrast.

4.7.1. Aloud > Silent Contrasts

The aloud > silent contrasts are of particular theoretical value, because they reveal apparent neural correlates of the production effect (defined as a behavioural contrast of aloud > silent). Our results here identified contributions of both reactivation and transformation, primarily in posterior portions of the left insula and precuneus respectively. The left insula has been linked to speech and language processes broadly (see Oh et al., 2014 for meta-analysis), while Woolnough et al. (2019) suggested that its posterior portion might mediate integration of somatosensory and auditory information during speech production. As such, reactivation in this area likely reflects reinstatement of speech-elicited processes, consistent with a distinctiveness-based account of the production effect (Forrin et al., 2012; Jamieson et al., 2016; MacLeod et al., 2010; MacLeod & Bodner, 2017).

Understanding the observed transformation effect in the precuneus is less straightforward, and requires careful consideration of the precuneus' known functions. This area has been associated with a number of processes, notably including episodic memory retrieval and metacognition. With respect to the former, fMRI work has implicated posterior precuneus in recollection (i.e., the subjective experience of remembering a specific event, as opposed to the more general sensation of familiarity) (Fandakova et al., 2021; Henson et al., 1999), memory for contextual details (e.g., source memory judgements; Lundstrom et al., 2003, 2005), and vivid reminiscence (Richter et al., 2016). The precuneus has also been linked to memory-relevant metacognition—that is, evaluation of one's internal cognitive state as it pertains to mnemonic decisions. One study reported that grey matter concentration in posterior precuneus was correlated with participants' ability to appropriately evaluate their own performance on a memory task

(selecting a previously studied stimulus from two alternatives; McCurdy et al., 2013), while self-evaluation during a similar task also elicited activation in the precuneus (albeit in a more inferior portion) relative to a perceptual judgement task (Morales et al., 2018). Moreover, a TMS study reported that selective disruption of the precuneus perturbed participants' ability to appropriately evaluate their own memory (Ye et al., 2018). In a slightly different vein, the paracingulate network—which is contained within the precuneus, and importantly includes its posterior portion (Dadario & Sughrue, 2023)—is uniquely suited to incorporate speech information as a component of both recollection and metacognition. This network receives input from multiple areas, notably including the pre/postcentral gyri, insular, and supplementary motor area; all of which are involved in speech production (e.g., Bailey et al., 2021). Dadario and Sughrue (2023) suggest that this network acts as a hub for integrating external sensory information (via functional connections to the aforementioned areas) with internal knowledge (e.g., introspective information and knowledge of task goals) for the purpose of guiding goal-directed behaviour. Therefore, we reason that the precuneus may have the capacity to incorporate encoded sensorimotor information (that was elicited during speech production) during memory-related metacognition, particularly in the context of a specific cognitive goal (e.g., deciding whether a presented word is old or new).

Based on the above information, we suggest that transformation in the precuneus may reflect access to, and subsequent metacognitive evaluation of encoded speech-related information, for the purpose of making old-new judgments during the recognition task. This encoded (and subsequently evaluated) information may be received as input from the insula, which is functionally connected to the paracingulate network (Dadario & Sughrue, 2023) and which exhibited reactivation effects in our analysis. Our proposal is consistent with the notion that participants' recognition judgements on aloud words are guided by an evaluative heuristic (MacLeod & Bodner, 2017); in this case, however, remembering is not entirely guided by faithful reinstatement ("replaying") of encoded experiences. As discussed in the introduction, evaluating encoded information, as opposed to mentally replaying it, may require reorganisation of that information

(consistent with prior multivariate studies showing goal-dependent changes in neural decodability, e.g., Favila et al., 2018; Nastase et al., 2017; Wang et al., 2018).

An alternative interpretation (though one which is not mutually exclusive with a heuristic account) is that transformation in the precuneus may reflect some component(s) of the retrieval process described by Wakeham-Lewis et al. (2022). By this account, when presented with a word (for which they must make an old/new decision), participants may simulate reading the presented word aloud and then compare the simulated experience with available encoded information associated with that word. It seems feasible that the precuneus might play a role here: in addition to its roles in recollection and metacognition, the precuneus is reportedly activated by a number of tasks involving mental simulation from a first-person perspective, such as mental imagery and imagined navigation (see Cavanna & Trimble, 2006 for review). More recently, Tanaka and Kirino (2021) report that during imagined singing, the precuneus exhibited heightened functional connectivity with perisylvian areas associated with speech production (inferior frontal and temporal gyri), as well as the middle and (medial) superior frontal gyri, which have both been linked to planning and cognitive control of speech output (Bourguignon, 2014; Hertrich et al., 2016, 2021). Takana and Kirino (2021) suggest that their results may reflect integration of language-related information during imagined singing. With respect to our study, because singing bears perceptual/experiential similarities to reading aloud (vocalisation and sensory feedback), it seems feasible that the precuneus would play a similar role in imagined aloud reading. Having said this, it remains unclear whether our observed transformation effect reflects simulation *per-se*, or the process of comparing the simulated experience with the encoded experience (Wakeham-Lewis et al., 2022). As discussed above, the metacognitive functional properties of the precuneus seem compatible with evaluation of encoded sensorimotor information; in this case, comparing mental simulation to past experiences. As before, such evaluation may be possible owing to the precuneus' functional connections with language-related areas involved in the initial productive experience. Considering that (by Wakeham-Lewis and colleagues' account) simulation should take place on *all* recognition trials (including both aloud and

silent words), our aloud > silent transformation effect may reflect successful matching between the simulated speech and encoded information on (correctly recognised) aloud trials. Unfortunately, it is difficult to separate mental simulation from comparison (to encoded information) based on the current data; we feel that further research is required to properly address this issue.

4.7.2. Silent > Aloud Contrasts

The silent > aloud reactivation contrast revealed a number of clusters; notably, clusters with the strongest evidence were situated in posterior fusiform and prefrontal cortices. Prior work has indicated that posterior fusiform cortex is sensitive to low-level orthographic properties of printed words (e.g., consonants versus false fonts) (Thesen et al., 2012; Vinckier et al., 2007); its role in our study likely reflects reinstatement of these visual features during successful recognition. This finding is consistent with previous multivariate work reporting reactivation in early visual areas during encoding and retrieval of images (e.g., Bone & Buchsbaum, 2021; Bosch et al., 2014). We note that an adjacent cluster in the left temperooccipital portion of the fusiform gyrus (Cluster #9 in Table 4.2) approximately corresponds to the visual word-form area (VWFA)[18], which has been the focus of much work surrounding the neural correlates of reading. The VWFA is generally considered central to comprehension of visual word-forms (e.g., Brem et al., 2010; Carreiras et al., 2014; Cohen et al., 2002; McCandliss et al., 2003; Petersen et al., 1990; Turkeltaub et al., 2014)—that is, perceptually-invariant mental representations of printed words (Warrington & Shallice, 1980). Our finding that this area (in addition to posterior fusiform gyrus) demonstrated reactivation for silent words may signal greater

[18] Multiple studies (e.g., L. Chen et al., 2019; Martin et al., 2015; Vogel et al., 2012) consider the putative VWFA to be centred on the following coordinates: x = −45, y = −57, z = −12. The COG for the cluster we identified in temperooccipital fusiform cortex was x = −42, y = −60, z = −15; within 3 voxels of putative VWFA in any direction.

dependence (relative to aloud words) on orthographic information during recognition, likely because the initial encoding experience was entirely visual.

With respect to prefrontal areas, prior work has implicated both the pars opercularis and supplementary motor area in processes necessary for speech planning and production. In the context of single word reading, the pars opercularis has traditionally been associated with grapheme-to-phoneme mapping (Fiez et al., 1999; Mechelli et al., 2005), while supplementary motor area is thought to play a role in cognitive control of speech-related motor processes such as initiation and timing(Hertrich et al., 2016). As for the superior frontal gyrus, this area has been linked to domain-general cognitive control, particularly cognitive flexibility (i.e., task- or response rule-switching) and response inhibition (e.g., Cutini et al., 2008; Konishi et al., 2003; see Niendam et al., 2012 for a meta-analysis of these processes). Given the appearance of reactivation in multiple areas that are either involved in speech planning or cognitive control, we suggest that prefrontal reactivation might reflect reinstatement of (item-specific) inhibitory 'codes' related to suppression of vocal responses during silent reading. These codes may facilitate a recognition heuristic similar to that described by Macleod and Bodner (2017), but in this case signalling that words were *not* produced (*I remember stopping myself from reading this word aloud, therefore I must have seen it before*).

4.7.3. Conclusion

This study was the first to investigate neural pattern reactivation and transformation during recognition in the context of the production effect. Our results broadly support a distinctiveness-based account whereby recognising aloud words depends on retrieval of item-specific speech information. That being said, this conclusion rather depends on how one conceptualises retrieval. Speech information may be reinstated in its original encoded format, consistent with the idea of mentally "replaying" sensorimotor experiences (MacLeod & Bodner, 2017). On the other hand, the encoded information may be manipulated or reorganised, perhaps reflecting metacognitive

evaluation of the recollected (or simulated; Wakeham-Lewis et al., 2022) experience. Recognising silent words seems relatively more dependent on reinstatement. This may be because, unlike aloud words, silent words were not encoded alongside the unique sensorimotor experiences associated with articulation; therefore, participants must rely more heavily on visual-orthographic information.

4.8. Data And Code Availability

Code for all analyses reported in this manuscript is publicly available on GitHub [1]; additional materials that are necessary for analyses are stored in an Open Science Framework (OSF) repository [2]. Raw data from a subset of participants (who consented to their anonymized data being made publicly available) is available on OSF [2]. Note that the data reported in this manuscript are from the "PE" experiment described in both repositories.

[1] https://github.com/lbailey25/Production_Effect_MVPA

[2] https://osf.io/czb26/?view_only=86a66caf1d71484d8ef0293cfa2371df

4.9. Acknowledgements

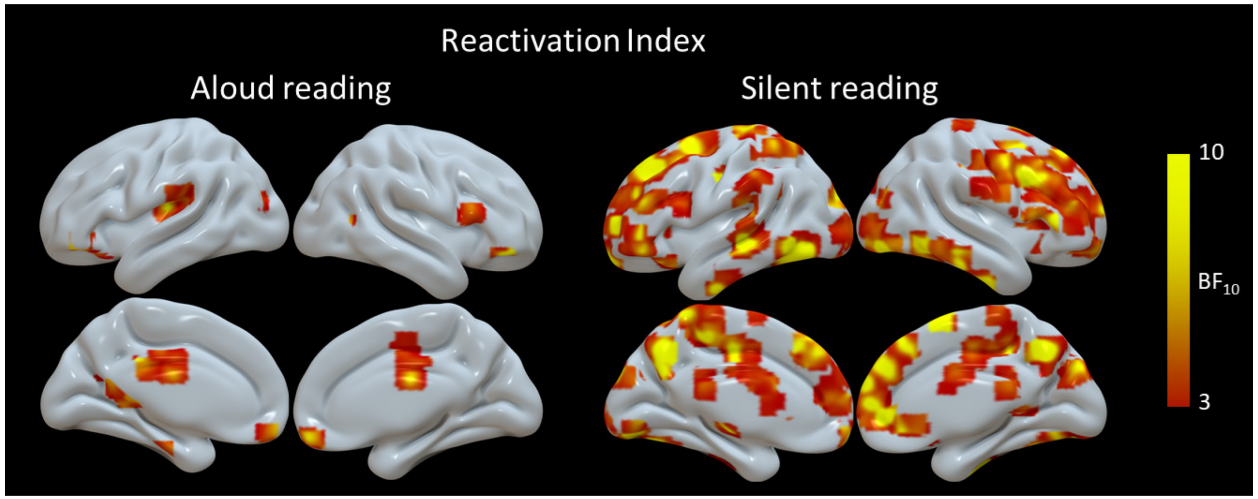
We wish to thank the following individuals for their assistance with data collection: Matt Rogers, Laura McMillan, Cindy Hamon-Hill. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to GEB and AJN (Grant numbers: RGPIN-2015-04131, RGPIN-2017-05340). LMB was supported by a Killam Predoctoral scholarship.

4.10. CRediT authorship statement

Lyam M. Bailey: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Project

administration, Funding acquisition. **Heath E. Matheson:** Conceptualization, Methodology, Software, Writing - review & editing. **Jonathan M. Fawcett:** Conceptualization, Writing - review & editing. **Glen E. Bodner:** Conceptualization, Funding acquisition. **Aaron J. Newman:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition.

4.11. Supplementary Materials



Supplementary Figure 4.1. Surfaces show BF_{10} maps for the reactivation analysis, separately for the read aloud condition (left) and read silently condition (right).

Supplementary Table 4.1. Clusters exhibiting evidence of reactivation in each reading condition. Anatomical labels in boldface reflect the structure in which the centre of mass (COM) for that cluster was located. Values in parentheses for each anatomical label indicate the number of searchlight centres (i.e., voxels) within that label contributing to the cluster.

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
Aloud	1	45	12.76	74.6	42	40	-17	**R Frontal Pole (45)**
	2	25	9.57	34.7	-35	38	-8	**L Frontal Pole (19)**
								L Frontal Orbital Cortex (6)
	3	25	9.36	42.4	1	54	-16	**R Frontal Medial Cortex (8)**

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
								L Frontal Medial Cortex (11)
								L Frontal Pole (1)
								R Frontal Pole (5)
	4	292	8.29	150	-35	-26	14	**L Heschls Gyrus (27)**
								L Central Opercular Cortex (15)
								L Insular Cortex (67)
								L Parietal Operculum Cortex (99)
								L Planum Polare (1)
								L Planum Temporale (25)
								L Postcentral Gyrus (2)
								L Superior Temporal Gyrus posterior (5)
								L Supramarginal Gyrus anterior (5)
								L Supramarginal Gyrus posterior (1)
								Left Putamen (2)
	5	30	8	41.9	-11	-31	27	**L Cingulate Gyrus posterior (16)**
	6	29	7.25	29.9	2	-13	24	**R Cingulate Gyrus posterior (9)**
								L Cingulate Gyrus anterior (4)
								L Cingulate Gyrus posterior (2)
								R Cingulate Gyrus anterior (4)
								Right Lateral Ventricle (3)
	7	32	7.22	20.4	-25	-63	28	**L Lateral Occipital Cortex superior (18)**
								L Precuneous Cortex (14)
	8	39	6.28	22.9	-15	-43	9	**L Cingulate Gyrus posterior (13)**
								L Precuneous Cortex (8)
	9	26	5.85	12.1	-23	-19	-30	**L Parahippocampal Gyrus anterior (26)**
	10	24	5.66	14.2	-40	24	-23	**L Temporal Pole (14)**
								L Frontal Orbital Cortex (10)

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
	11	27	5.34	9.11	-30	-78	18	**L Lateral Occipital Cortex superior (27)**
	12	50	5.26	16.7	1	-15	38	**R Cingulate Gyrus anterior (17)**
								L Cingulate Gyrus anterior (11)
								L Cingulate Gyrus posterior (19)
								R Cingulate Gyrus posterior (3)
	13	25	5.03	10.1	40	-55	9	**R Middle Temporal Gyrus temporooccipital (23)**
	14	20	4.89	12	53	18	15	**R Inferior Frontal Gyrus pars opercularis (20)**
Silent	1	71	122.4	4410	48	-18	-35	**R Inferior Temporal Gyrus posterior (71)**
	2	409	48.85	2290	-42	-66	-11	**L Occipital Fusiform Gyrus (86)**
								L Inferior Temporal Gyrus temporooccipital (70)
								L Lateral Occipital Cortex inferior (144)
								L Middle Temporal Gyrus temporooccipital (25)
								L Temporal Occipital Fusiform Cortex (84)
	3	258	25.33	447	0	-61	42	**R Precuneous Cortex (94)**
								L Precuneous Cortex (164)
	4	38	22.05	279	-42	-11	36	**L Precentral Gyrus (34)**
								L Postcentral Gyrus (4)
	5	526	21.74	680	-18	30	44	**L Superior Frontal Gyrus (285)**
								L Frontal Pole (82)
								L Middle Frontal Gyrus (157)
								L Precentral Gyrus (2)
	6	100	16.93	544	22	5	50	**R Superior Frontal Gyrus (61)**
								R Middle Frontal Gyrus (25)
								R Precentral Gyrus (14)
	7	39	14.49	94.5	-55	-10	-39	**L Inferior Temporal Gyrus anterior (27)**
								L Inferior Temporal Gyrus posterior (12)
	8	38	14.3	118	-47	39	23	**L Frontal Pole (20)**

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
								L Middle Frontal Gyrus (18)
	9	29	13.31	53.7	9	8	63	**R Superior Frontal Gyrus (19)**
								R Supplementary Motor Area (10)
	10	77	13.15	217	59	18	8	**R Inferior Frontal Gyrus pars opercularis (37)**
								R Inferior Frontal Gyrus pars triangularis (29)
								R Precentral Gyrus (11)
	11	77	13.03	188	-47	-20	-22	**L Inferior Temporal Gyrus posterior (75)**
								L Middle Temporal Gyrus posterior (2)
	12	92	12.38	201	36	-85	-11	**R Lateral Occipital Cortex inferior (74)**
								R Occipital Fusiform Gyrus (10)
								R Occipital Pole (8)
	13	114	12.14	200	-61	-30	-5	**L Middle Temporal Gyrus posterior (95)**
								L Superior Temporal Gyrus posterior (19)
	14	111	12.08	181	46	-58	-12	**R Inferior Temporal Gyrus temporooccipital (46)**
								R Lateral Occipital Cortex inferior (12)
								R Occipital Fusiform Gyrus (12)
								R Temporal Occipital Fusiform Cortex (41)
	15	74	11.99	88	-26	57	7	**L Frontal Pole (74)**
	16	22	11.07	57.3	-49	46	2	**L Frontal Pole (22)**
	17	157	10.72	85.3	-14	-10	47	**L Supplementary Motor Area (67)**
								L Cingulate Gyrus anterior (15)
								L Cingulate Gyrus posterior (1)
								L Precentral Gyrus (10)
								L Superior Frontal Gyrus (13)
	18	120	10.59	100	64	-37	-20	**R Inferior Temporal Gyrus posterior (41)**
								R Inferior Temporal Gyrus temporooccipital (50)
								R Middle Temporal Gyrus posterior (14)

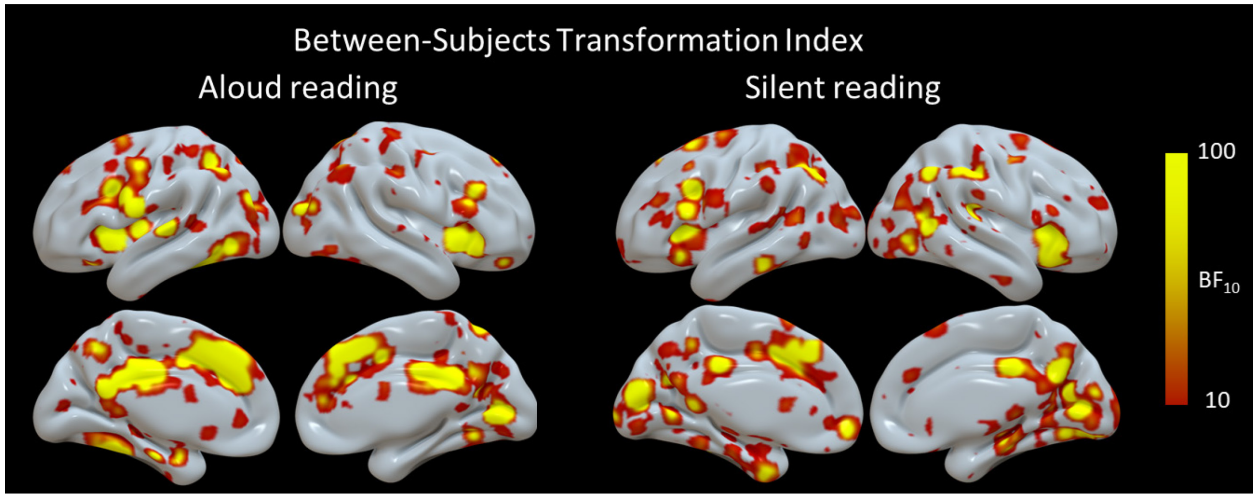
Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
								R Middle Temporal Gyrus temporooccipital (15)
	19	36	10.58	45	60	-26	49	**R Supramarginal Gyrus anterior (30)**
								R Postcentral Gyrus (6)
	20	117	10.49	48.7	-13	62	-17	**L Frontal Pole (117)**
	21	108	10.29	107	47	-10	28	**R Postcentral Gyrus (74)**
								R Central Opercular Cortex (1)
								R Precentral Gyrus (25)
								R Supramarginal Gyrus anterior (2)
	22	786	10.29	475	30	22	37	**R Middle Frontal Gyrus (423)**
								R Frontal Pole (47)
								R Inferior Frontal Gyrus pars opercularis (22)
								R Inferior Frontal Gyrus pars triangularis (2)
								R Paracingulate Gyrus (1)
								R Precentral Gyrus (91)
								R Superior Frontal Gyrus (116)
								Right Caudate (9)
	23	27	9.94	33.7	31	-79	21	**R Lateral Occipital Cortex superior (27)**
	24	142	9.83	59.9	-14	-31	60	**L Precentral Gyrus (84)**
								L Postcentral Gyrus (58)
	25	313	9.57	122	-22	-86	-4	**L Occipital Fusiform Gyrus (128)**
								L Intracalcarine Cortex (1)
								L Lateral Occipital Cortex inferior (35)
								L Lingual Gyrus (19)
								L Occipital Pole (130)
	26	422	9.3	94.1	12	49	-1	**R Paracingulate Gyrus (85)**
								L Frontal Pole (11)
								L Paracingulate Gyrus (77)

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
								R Cingulate Gyrus anterior (33)
								R Frontal Medial Cortex (24)
								R Frontal Pole (158)
								R Subcallosal Cortex (21)
	27	236	9.11	47.7	6	53	29	**R Superior Frontal Gyrus (59)**
								L Frontal Pole (27)
								L Superior Frontal Gyrus (5)
								R Frontal Pole (104)
								R Paracingulate Gyrus (41)
	28	188	8.68	166	-23	46	27	**L Frontal Pole (179)**
								L Middle Frontal Gyrus (1)
								L Superior Frontal Gyrus (8)
	29	29	8.38	65	-32	26	-22	**L Frontal Orbital Cortex (29)**
	30	201	8.27	93	-22	4	65	**L Superior Frontal Gyrus (180)**
								L Middle Frontal Gyrus (19)
								L Precentral Gyrus (2)
	31	20	8.05	24.9	-55	-30	19	**L Parietal Operculum Cortex (20)**
	32	44	7.48	54.8	-6	-44	70	**L Postcentral Gyrus (42)**
								L Precuneous Cortex (2)
	33	53	7.44	26.6	10	-80	29	**R Cuneal Cortex (51)**
								R Lateral Occipital Cortex superior (2)
	34	54	7.37	36.6	-15	-90	27	**L Occipital Pole (38)**
								L Cuneal Cortex (10)
								L Lateral Occipital Cortex superior (6)
	35	34	7.17	23.3	35	15	16	**R Frontal Operculum Cortex (32)**
								R Inferior Frontal Gyrus pars opercularis (2)
	36	70	7.02	27.3	-25	-55	60	**L Superior Parietal Lobule (64)**

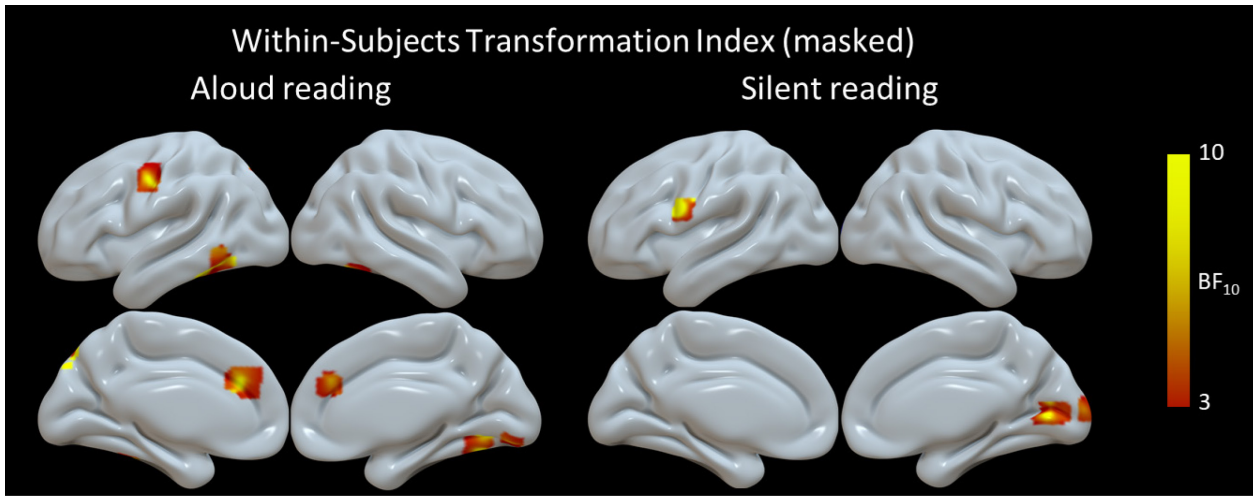
Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
								L Lateral Occipital Cortex superior (6)
	37	34	6.94	28.1	-48	40	-9	**L Frontal Pole (29)**
								L Frontal Orbital Cortex (2)
								L Inferior Frontal Gyrus pars triangularis (3)
	38	204	6.86	37.5	29	51	14	**R Frontal Pole (202)**
								R Middle Frontal Gyrus (2)
	39	38	6.84	22.3	-24	25	5	**L Frontal Orbital Cortex (7)**
								L Inferior Frontal Gyrus pars triangularis (1)
								L Insular Cortex (12)
	40	80	6.65	26	17	-32	51	**R Precentral Gyrus (43)**
								R Postcentral Gyrus (36)
								R Precuneous Cortex (1)
	41	40	6.53	23.2	-42	13	19	**L Inferior Frontal Gyrus pars opercularis (38)**
								L Precentral Gyrus (2)
	42	21	6.3	12.4	35	-85	-1	**R Lateral Occipital Cortex inferior (21)**
	43	91	6.19	27.3	1	-77	46	**R Precuneous Cortex (39)**
								L Precuneous Cortex (41)
								R Cuneal Cortex (10)
								R Lateral Occipital Cortex superior (1)
	44	74	6.11	15.8	20	22	-10	**R Frontal Orbital Cortex (57)**
								R Insular Cortex (2)
								R Subcallosal Cortex (8)
	45	26	5.95	14.3	32	-96	4	**R Occipital Pole (26)**
	46	65	5.87	30.8	-64	-35	41	**L Supramarginal Gyrus anterior (58)**
								L Supramarginal Gyrus posterior (7)
	47	27	5.87	12.5	-52	-30	7	**L Planum Temporale (26)**
								L Superior Temporal Gyrus posterior (1)

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
	48	69	5.77	27.1	10	-12	38	**R Cingulate Gyrus anterior (26)**
								R Cingulate Gyrus posterior (11)
								R Precentral Gyrus (3)
								R Supplementary Motor Area (7)
								Right Lateral Ventricle (1)
	49	62	5.7	21.2	-39	-40	28	**L Parietal Operculum Cortex (32)**
								L Superior Temporal Gyrus posterior (3)
								L Supramarginal Gyrus anterior (1)
								L Supramarginal Gyrus posterior (16)
	50	41	5.53	12.3	-49	-41	56	**L Supramarginal Gyrus posterior (19)**
								L Postcentral Gyrus (11)
								L Superior Parietal Lobule (9)
								L Supramarginal Gyrus anterior (2)
	51	29	5.53	14.1	-29	-12	51	**L Precentral Gyrus (29)**
	52	37	5.38	12.9	36	16	8	**R Frontal Operculum Cortex (23)**
								R Insular Cortex (14)
	53	26	5.35	12.8	-43	52	-12	**L Frontal Pole (26)**
	54	22	4.95	8.84	-30	-57	36	**L Lateral Occipital Cortex superior (11)**
								L Angular Gyrus (1)
								L Superior Parietal Lobule (9)
								L Supramarginal Gyrus posterior (1)
	55	29	4.84	13.2	1	-29	69	**R Precentral Gyrus (12)**
								L Precentral Gyrus (17)
	56	32	4.81	11.5	4	-3	29	**R Cingulate Gyrus anterior (27)**
								L Cingulate Gyrus anterior (5)
	57	26	4.74	8.7	52	20	25	**R Inferior Frontal Gyrus pars opercularis (24)**
								R Inferior Frontal Gyrus pars triangularis (1)

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N Voxels)
								R Middle Frontal Gyrus (1)
	58	20	4.34	7.31	49	45	3	**R Frontal Pole (20)**
	59	26	4.13	6.81	20	-30	69	**R Postcentral Gyrus (17)**
								R Precentral Gyrus (9)



Supplementary Figure 4.2. Surfaces show BF_{10} maps for the between-subjects transformation analysis, separately for the read aloud condition (left) and read silently condition (right). Note these maps have been thresholded at $BF_{10} \geq 10.0$ ($BF_{10} \geq 3.0$ was used for all other figures), so that areas exhibiting strong evidence are more easily identifiable.



Supplementary Figure 4.3. Surfaces show BF_{10} maps for the within-subjects transformation analysis, separately for the read aloud condition (left) and read silently condition (right). Each BF_{10} map has been masked with the that of the between-subjects analysis from the corresponding condition.

Supplementary Table 4.2. Clusters exhibiting evidence of within-subjects transformation (masked with between-subjects transformation) in each reading condition. Anatomical labels in boldface reflect the structure in which the centre of mass (COM) for that cluster was located. Values in parentheses for each anatomical label indicate the number of searchlight centres (i.e., voxels) within that label contributing to the cluster.

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
Aloud	1	69	72.34	766	-7	-83	44	**L Precuneous Cortex (17)**
								L Cuneal Cortex (7)
								L Lateral Occipital Cortex superior (43)

Condition	Cluster #	Cluster size (mm^3)	Mean BF_{10}	Max BF_{10}	x	y	z	Anatomical Label(s) (N voxels)
								R Cuneal Cortex (2)
	2	31	17.17	84.6	-46	-42	-18	**L Inferior Temporal Gyrus posterior (13)**
								L Inferior Temporal Gyrus temporooccipital (3)
								L Temporal Fusiform Cortex posterior (14)
								L Temporal Occipital Fusiform Cortex (1)
	3	103	9.32	54.1	-8	35	28	**L Paracingulate Gyrus (81)**
								L Cingulate Gyrus anterior (5)
								L Frontal Pole (1)
								L Superior Frontal Gyrus (4)
								R Paracingulate Gyrus (12)
	4	20	9.22	60.4	18	-82	-9	**R Occipital Fusiform Gyrus (13)**
								R Lingual Gyrus (7)
	5	53	8.24	46.6	32	-59	-11	**R Temporal Occipital Fusiform Cortex (39)**
								R Lingual Gyrus (4)
								R Occipital Fusiform Gyrus (10)
	6	21	7.45	48.1	-51	-56	-14	**L Inferior Temporal Gyrus temporooccipital (18)**
								L Middle Temporal Gyrus temporooccipital (3)
	7	21	7.19	23.6	-51	-8	43	**L Precentral Gyrus (21)**
Silent	1	36	11.12	81	-52	2	22	**L Precentral Gyrus (36)**
	2	44	9.57	33.7	6	-72	6	**R Intracalcarine Cortex (38)**
								R Lingual Gyrus (6)
	3	26	5.97	14	8	-100	9	**R Occipital Pole (26)**

CHAPTER 5. GENERAL DISCUSSION

5.1. Overview

The goal of this dissertation was to investigate neural correlates of the production effect, which is a behaviourally-defined memory advantage for words read aloud compared to words read silently (MacLeod et al., 2010). To address this goal, I have presented results from univariate and multivariate analysis of data from three fMRI experiments which entailed different adaptations of the standard production effect paradigm. Each experiment may be thought of as providing a different perspective on neural processes underlying the production effect. In brief, Chapter 2 asked the question: *which brain areas are involved in encoding and retrieval of words read aloud compared with words read silently?* Chapter 3 asked: *how is information about stimulus properties represented in activation patterns elicited by reading aloud compared with reading silently?* Chapter 4 asked: *how do different retrieval mechanisms—neural pattern reactivation and transformation—contribute to recognition of words read aloud compared with words read silently?*

Below I provide a brief summary of my findings and conclusions from each chapter (Section 5.2), followed by some discussion of methodological considerations (and their implications) in my experiments (Section 5.3). Following this, I discuss the implications of my findings in the context of the production effect literature (Section 5.4), as well as implications for broader research surrounding multivariate correlates of episodic memory (Section 5.5). Finally, I present some possible avenues for future research (Section 5.6), and some concluding remarks on the nature of the production effect (Section 5.7).

5.2. Review of Empirical Findings and Conclusions

5.2.1. Chapter 2

In the experiment for Chapter 2, participants completed a study phase (encoding) and a test phase (retrieval) while they underwent fMRI scanning. During the study phase,

participants were presented with a single word on each trial, and were instructed to read each word depending on a preceding cue—either read aloud, read silently, or read silently and say the word "check" (heron in, I will refer to this latter condition as the sensorimotor control condition). During the test phase, participants were presented with the same words from the study phase, intermixed with a set of previously unseen foil words, and were instructed to indicate their memory for each word using a remember/know paradigm (Tulving, 1985). I used univariate analyses to compare activation elicited in each condition during each phase. Results from the study phase revealed a number of brain areas that were preferentially activated by reading aloud compared to the silent and sensorimotor control conditions. Broadly, I found that reading aloud elicited preferential activation of frontal, temporal, and parietal areas associated with speech production, somatosensation, and audition. Moreover, activation in a subset of these areas (broadly comprising left inferior prefrontal cortex) was correlated with participants' recollective performance at test. Based on these findings, I concluded that distinctive sensorimotor processes that are involved in speech production likely play a role in encoding words that were read aloud.

Univariate analysis of test phase data revealed that inferior frontal, superior temporal, and ventral temporal areas (linked to motor control, auditory processing, and orthographic processing respectively) were preferentially activated by words from the aloud condition relative to the silent condition. Importantly, the test phase analysis was constrained to voxels that were activated by the aloud condition during the study phase, meaning that these areas were consistently activated by the reading aloud condition across encoding and retrieval. I therefore interpreted these results to reflect mental reinstatement of sensory / perceptual processes that had been engaged by the initial encoding experience.

Chapter 2 provided a good starting point for understanding neural correlates of the production effect, in that it revealed brain areas that are broadly involved in encoding and retrieval of words read aloud. Building upon these findings, Chapters 3 and 4 sought

to understand the information content of activation patterns elicited by aloud and silent reading, using multivariate analyses.

5.2.2. Chapter 3

The experiment for Chapter 3 was similar to the study phase from Chapter 2, but included some procedural changes designed to optimise scan time and to ensure ideal conditions for multivariate analysis. The major changes were that this experiment did not include a sensorimotor control condition, and that each stimulus was now presented four times instead of once (but see Section 3 for a discussion of other procedural changes). For this chapter, I used representational similarity analysis (RSA) to decode information concerning five stimulus properties (visual, orthographic, phonological, semantic, and articulatory) from activation patterns elicited by aloud and silent reading. My results revealed task-dependent variability in stimulus decodability. Aloud reading (relative to silent reading) was associated with greater decodability of visual, phonological, semantic, and articulatory information, mainly in frontal and parietal areas associated with speech planning and cognitive control of speech output. I attributed these increases in decodability to the cognitive demands of reading aloud: the individual must plan, execute, and monitor an appropriate sequence of articulatory movements to correctly produce each word. Information that is necessary for these cognitive operations (mainly, I argued, phonological and articulatory information) is therefore up-regulated in associated cortical areas. I also suggested that increased decodability of low-level visual information may reflect increased attention to words read aloud. By contrast, orthographic information was more decodable during silent reading (relative to aloud reading) in right anterior ventral temporal cortex. I suggested that this finding may reflect participants relying more heavily on phonologically-mediated access to semantics in the aloud reading condition, resulting in relatively more print-to-meaning mapping in the silent condition.

Chapter 3 went beyond Chapter 2 in two major ways. First, rather than simply identifying particular regions as being involved in and/or contributing to reading aloud,

with RSA I was able to examine the presence of specific types of information (articulatory, phonological, etc.). Second, owing to the fact that RSA is concerned with the complex response patterns elicited by unique stimuli, I was able to attribute my findings to stimulus-specific processing — as opposed to nonspecific processes that are elicited by all stimuli in a given reading condition. An important caveat here is that it is difficult to link my RSA results to the production effect directly, since this experiment did not have associated behavioural data (therefore it was not possible to link decodability to memory performance)[19]. That being said, and as I argued in Chapter 3, this experiment was largely representative of conditions which reliably give rise to the production effect, and so my results are likely informative as to its underlying encoding processes (but see Section 3 of this chapter for a more thorough discussion of this point). At minimum, however, we can conclude that the various types of information that were up-regulated by aloud reading (as revealed by my results) are at least *available* for participants to encode. This is a major assumption of many cognitive accounts of the production effect; providing evidence for this claim should therefore constitute progress in and of itself.

5.2.3. Chapter 4

The experiment for Chapter 4 was similar to that of Chapter 2 (but see Section 3 for a discussion of procedural differences). As before, participants underwent separate study and test phases (labelled "encoding" and "recognition" respectively in Chapter 4, in the interests of more transparent nomenclature). In the encoding phase participants read presented words either aloud or silently; during the recognition phase participants' memory was tested using an old/new judgement task. Here, I applied multivariate analyses to investigate behaviourally-relevant neural reactivation and transformation in the aloud and silent reading conditions. I quantified neural reactivation as the similarity (correlation) of response patterns to the same word between encoding and retrieval. This

[19] Experiment 3 did not test participants' memory because repeating each stimulus four times (required for RSA) was expected to yield ceiling recognition performance; see Section 3 for elaboration.

measure was behaviourally relevant in that it captured differences in these similarity values between correctly remembered and forgotten items (as per Zeithamova et al., 2017). I quantified neural transformation as the difference between (i) same-item similarity between separate encoding presentations and (ii) same-item similarity between encoding and recognition. As with the reactivation measure, transformation was computed as an effect size between remembered and forgotten items. Importantly, to ensure that I had captured true neural transformation (as opposed to weakened activation patterns or increased noise between encoding and retrieval), the transformation results were constrained using a second between-subjects transformation analysis (J. Chen et al., 2017). In brief, this between-subjects analysis identified areas where item-specific information could still be decoded at retrieval, despite the patterns having been altered between encoding and retrieval.

My results in Chapter 4 revealed different retrieval dynamics depending on whether words had been read aloud or silently at encoding. Most notably, recognition of aloud words (compared to silent words) was associated with neural transformation in the posterior portion of the left precuneus. I attributed this finding to recollection and subsequent metacognitive evaluation of encoded sensorimotor experiences. I drew these conclusions mainly based on the precuneus' known role in detailed recollection and metacognition, as well as its functional connections to areas responsible for sensorimotor processing. By contrast, recognising words that were read silently seemed to depend more heavily on reactivation, mainly in prefrontal areas associated with cognitive control and posterior ventral temporal areas associated with orthographic analysis. Taken together, these two sets of results suggest that remembering involves variable contributions of different neural mechanisms (reactivation and transformation), depending on the preceding encoding conditions.

Chapter 4 built upon the previous two chapters in two ways. Unlike Chapter 2, the use of multivariate analysis allowed me to investigate processes underlying recognition of individual words, as opposed to all words in each condition. Moreover, unlike Chapter

3, my analyses for Chapter 4 were linked to participants' recognition performance, meaning that my findings are directly relevant to the behavioural production effect.

5.3. Deviations From The Standard Production Effect Paradigm

Before I attempt to contextualise my findings and conclusions within the wider production effect literature, it is important to note that all three of my experiments deviated methodologically (to varying degrees) from the classic production effect paradigm. It is possible that these deviations (i) might have led to changes in how participants processed the stimuli, particularly during encoding, or (ii) otherwise place constraints on interpretation of my results. I will address these potential concerns here.

In all three experiments, participants read each word according to an instructional cue that was presented before the word was presented; this differs from prior work in which words and instructions were presented concurrently (e.g., in MacLeod et al., 2010 participants were instructed to read each word according to the colour in which it was presented). I do not think it likely that this change fundamentally undermined the cognitive processes giving rise to the production effect, given that the experiments in Chapters 2 and 4 both elicited the expected recognition advantage for aloud words. Moreover, at least one other study has elicited a production effect using this instructional procedure (Hassall et al., 2016); while a more recent study reported a production effect when reading instructions were separated into different blocks (B. Zhang et al., 2023). That being said, this procedure places constraints on interpretation of the fMRI results. In all of my analyses, the BOLD response for each trial was estimated based on the time window in which the stimulus word was presented; this means that all other time points within the trial (including the period in which the instructional cue was presented) were relegated to baseline. As such, any variance in the BOLD signal that was induced by different trial instructions (e.g., differences between "read aloud" and "read silently") would have ended up in the error term. The result of this is that my analyses would have been blind to neural activation underlying stimulus-independent, task-directed

cognition—that is, planning to name *an upcoming* word aloud (or silently, etc.). To be clear, this does not preclude activation associated with reading (according to task instruction) *the specific word* presented on each trial; it simply means that any activation associated with preparatory cognition ("I must be ready to read the upcoming word aloud / silently") would be ignored.

In the experiments from Chapters 3 and 4, participants read each word multiple times: either four times (Chapter 3) or twice (encoding phase of Chapter 4). I designed the experiments in this way to provide optimal conditions for multivariate analysis (Zeithamova et al., 2017). This design change may have affected cognitive processing relevant to the production effect. To my knowledge, only one study on the production effect has manipulated stimulus repetition: Ozubko et al. (2014) presented silent words (but not aloud words) to participants multiple times with the aim of manipulating encoding strength for those items independently of distinctiveness (consistent with the more general finding that repetition improves subsequent memory; Hintzman, 1976). While not the main focus of the experiment, the authors noted that this procedure made it possible to abolish the production effect, because recognition performance on (strengthened) silent items became equivalent to that of aloud items (Ozubko et al., 2014; Experiment 2). Beyond this study, to my knowledge no work has investigated effects of repeating presentations in *both* conditions. Considering that repetition should strengthen items, one might predict that it should improve participant's global recognition performance (regardless of condition), resulting in a smaller possible range for between-condition differences as performance approaches ceiling in both conditions. Indeed, during early piloting for this work, I found that exceeding two presentations at encoding abolished the production effect entirely because (pilot) participants reliably achieved ceiling performance in both conditions.

A more interesting question is whether repetition might *differentially* affect recognition for aloud and silent items. This question may be addressed by comparing results from Chapter 2 with those of Chapter 4 (in the experiment for the former,

participants saw each word only once during encoding). Mean recognition performance was similar across chapters for aloud words (79% in Chapter 2 and 81% in Chapter 4; an independent-samples Bayes *t* test revealed non-credible evidence for a difference between chapters, $BF_{10} = 0.33$), while it increased for silent words (60% in Chapter 2 and 74% in Chapter 4, with strong evidence for a difference between chapters, $BF_{10} = 24.72$). This suggests that silent words benefit more from repetition than do aloud words[20]. It seems, therefore, that my experiments gave rise to a disproportionate boost to encoding strength for silent items only (compared with conditions in which each word is presented only once). This effect may have masked processing advantages afforded to aloud words (relative to silent) that are present in the typical production effect paradigm.

The experiments from Chapters 3 and 4 also involved an active baseline task: between trials, participants made odd / even judgements on a series of presented numbers. I implemented this active baseline task because (i) it has been shown to improve signal-to-noise ratio in fMRI compared with the conventional resting baseline (Stark & Squire, 2001), and (ii) it ensured that participants remained engaged in the task despite very long periods between stimuli (particularly the Chapter 4 experiment, which entailed a stimulus onset asynchrony of 12 seconds). It is possible that inserting this task between trials might have affected encoding-related cognition during these experiments. Jonker et al. (2014) suggested that under pure list conditions (e.g., in between-subjects manipulations), participants may encode relational (item order) information for silent items. These authors reason that the mixed-list (within-subjects) production effect may therefore be driven, at least in part, by aloud items disrupting this relational processing, resulting in poorer encoding of silent items. One may argue that the insertion of an active baseline task might have further disrupted relational processing in my mixed-list experiments, leading to a greater disadvantage for silent items. While certainly possible, I suggest that such an effect would have had only minimal bearing on my results.

[20] As an aside, improved recognition performance in the silent condition is unlikely to be driven by other methodological factors such as the introduction of an active baseline task (discussed in the next paragraph).

Relational processing should already be disrupted in a mixed list design, meaning that any effect of the active baseline task would be additive (rather than introducing an entirely new form of perturbation). It is also worth noting that Jonker and colleagues' account is exclusively concerned with free-recall memory, and is therefore not applicable to recognition memory. True, some degree of relational processing may be intrinsic to encoding, regardless of the format of the subsequent memory test. Even so, Jonker and colleagues point out that participants' predictions about the upcoming test format may influence whether (or the extent to which) they encode relational information. Since my participants practised encoding and retrieval phases before beginning the experiment(s) proper, they would have expected to be tested on recognition memory (rather than free recall); it therefore seems unlikely that they would have invested in a relational encoding strategy to begin with. Nevertheless, I concede that participants may have engaged in some degree of relational processing (however minimal) when encoding silent items, and that the active baseline task may have disrupted this processing to a greater extent than would ordinarily occur in a standard mixed list design. In this sense, Experiments 3 and 4 may not have been fully representative of the cognitive conditions which typically give rise to the production effect (but see below).

In summary, my experiments entailed three key deviations from the conventional production effect paradigm: participants saw preceding (as opposed to accompanying) instructional cues, stimuli were repeated multiple times (Chapters 3 & 4 only), and participants also completed an active baseline task between trials (Chapters 3 & 4 only). The use of preceding instructional cues likely restricted my results to stimulus-specific processes (e.g., "I must read aloud the specific word that is being presented to me"), as opposed to more general goal-related cognition (e.g., "I must be ready to read the upcoming word aloud"). Based on inspection of condition-wise means from Chapters 2 and 4, I noted that stimulus repetition may have conferred a disproportionate advantage to silent words; this may have diminished neural effects shown by aloud > silent contrasts. Finally, the use of an active baseline task may have disrupted relational processing for silent items to a greater extent than would ordinarily occur in mixed-list conditions. Again,

this may have induced some changes to cognition during encoding. All of this being said, the above considerations should not be unduly over-emphasised. It is important to note that, despite these methodological irregularities, I still observed the expected recognition advantage for aloud items in Chapters 2 and 4 (participants' memory was not tested in Chapter 3), indicating that whichever processes give rise to the production effect were at least partially preserved. Moreover, an oft-cited feature of the production effect is its generalisability (i.e., robustness in the face of methodological deviations)—as discussed in Chapter 1, the production effect has been elicited using various stimulus types (words, pictures, text passages), forms of production (reading aloud, mouthing, singing, drawing), factorial designs (within- and between-subjects) and instructional methods (accompanying or preceding cues, blocked instructions). Viewed in this light, the methodological changes in my experiments easily fall within the range of contexts in which the production effect has been observed. As such, I do not feel that these changes fundamentally undermine the interpretation of my results, broadly, as neural correlates of the production effect.

5.4. Implications For Cognitive Accounts Of The Production Effect

In this section I will attempt to contextualise my findings within, and comment on their implications upon the wider production effect literature. While the various theoretical accounts of the production effect have been based almost exclusively on behavioural data, they nevertheless provide useful frameworks through which to interpret my neuroimaging results. In many respects, my results seem compatible, at least to some extent, with each of the cognitive accounts laid out in Chapter 1, as described below.

5.4.1. Agreement With Prior Accounts

First I will address relative distinctiveness, since this account rather dominates the literature surrounding the production effect. A central tenet of this accounts is that

reading aloud elicits sensorimotor experiences (proprioceptive feedback from the tongue and jaw, vibration of the larynx, auditory feedback) which are appended as features to the encoded record for having seen that word (often referred to as the *memory trace*). These additional features are thought to make the memory trace for aloud words distinctive (in that they stand out against the backdrop of non-distinctive silent items; MacLeod et al., 2010), which in turn facilitates subsequent retrieval of aloud words (Forrin et al., 2012; Jamieson et al., 2016; MacLeod et al., 2010; MacLeod & Bodner, 2017). Based on this account, one should expect encoding of aloud words to depend, to some extent, on speech-related neural processing, since speech processes are thought to give rise to privileged encoding. This prediction was borne out most clearly by Chapter 2, in which I reported that sensorimotor and auditory cortices were preferentially activated by the aloud reading condition compared with silent reading and the sensorimotor control condition. Moreover, in Chapter 3 I identified greater decodability of articulatory and phonological information during aloud relative to silent reading. As I argued in that chapter, both of these kinds of information are necessary for speech production; therefore I suggest that this set of findings fits with a relative distinctiveness account by which speech-elicited information is incorporated into the encoded memory trace. As an aside: the strongest evidence for phonological and articulatory information in Chapter 3 was located in medial frontal areas responsible for speech planning and cognitive control. Curiously, prior literature on relative distinctiveness makes no reference to speech planning or cognitive control, despite these being necessary components of speech production. Based on my findings, I suggest that proponents of this account should slightly modify their descriptions of encoding to incorporate such executive processes, in addition to sensorimotor and auditory feedback, into the distinctive memory trace.

Strength-based accounts concerning semantic processing and attention propose that aloud words are encoded more effectively than silent words, rather than being more distinctive. With respect to the semantic account, Fawcett et al. (2022) proposed that aloud words benefit from enhanced semantic processing. More precisely, these authors suggest that reading aloud facilitates integration of semantic features into the encoded

memory trace. From this account, one might expect aloud reading to involve preferential activation in cortices linked to semantic processing, or greater decodability of semantic information, relative to silent reading. Again, my results seem in line with these predictions. In Chapter 2, I found that reading aloud preferentially activated left superior temporal and inferior frontal areas (both of which have been implicated in semantic processing; Rodd et al., 2015; Vigneau et al., 2006), while in Chapter 3 I found that aloud reading was associated with greater decodability of semantic information compared with silent reading. These findings provide neural evidence for a role of privileged semantic processing when reading aloud.

Attention has also been posed to play a role in the production effect (Fawcett & Ozubko, 2016; Mama et al., 2018; Mama & Icht, 2019; Ozubko, Gopie, et al., 2012). Since aloud trials require an overt response (whereas silent trials do not), it is possible that participants allocate more attention to aloud words, and therefore encode those words more effectively. Importantly, this account does not specify attention to any particular stimulus features, but rather a *global* increase in cognitive investment when presented with words on aloud trials. As such, throughout this dissertation, I operationalised the attentional account as increased weighting on (i.e., greater decodability of) all stimulus features during aloud reading. From this perspective, the results of Chapter 3, which entailed greater decodability of multiple types of information during aloud reading, are arguably in line with an attentional account. True, up-regulation of phonological, articulatory, and semantic information may be explained by distinctiveness and semantic accounts detailed earlier. However, my results concerning visual information—which was more decodable in the aloud condition—is of particular interest here. In Chapter 3 I argued that, unlike phonological and articulatory information, visual information is no more relevant to aloud reading than it is to silent reading (that is to say, visual information is essential for both tasks). My finding that visual information is up-regulated during aloud reading, therefore, indicates that participants may be investing more cognitive resources in processing words on aloud trials.

Each of the accounts described above are primarily concerned with encoding processes; common to all of them is the notion that, for one reason or another, aloud words are encoded more effectively than silent words. My results from Chapters 2 and 3 (which contrasted encoding processes between aloud and silent reading) are therefore directly relevant to these accounts. As argued above, many of my findings are compatible with relative distinctiveness, enhanced semantic processing, and increased attention. One caveat here is that the results from Chapter 3 arguably provide more credible support for a role of distinctiveness, compared with semantics and attention. More precisely, clusters exhibiting evidence for semantic and visual decodability (from the aloud > silent contrasts) were of smaller spatial extent, and generally lower quantitative strength of evidence, compared with the phonological and articulatory results. For example: the strongest evidence (as quantified by Bayes factors) for semantic decodability came from a cluster with Mean, Max BF_{10} = 8.58, 24.60 (I observed similar values for the strongest visual cluster); this is in stark contrast to equivalent values for phonological (172.96, 882.0) and articulatory (36.71, 294) decodability. This is not to say that the evidence for roles of semantics and attention was not credible; it simply means that the strongest evidence available (in this study) seems to favour a distinctiveness-based account.

Other accounts are concerned with retrieval processes. One such account holds that participants may recollect the productive experience (complete with constituent sensorimotor and auditory experiences) associated with each word during a recognition task; successful mental reinstatement of the episode allows participants to endorse or reject the presented word using a heuristic: *I remember reading this word aloud, therefore I must have studied it* (MacLeod & Bodner, 2017)[21]. Presumably, such recollection should entail reinstatement of speech-specific neural states that were

[21] Similarly to the relative distinctiveness account, this account emphasises the importance of encoded speech-related features; indeed, it is often labelled a "distinctiveness heuristic" and presented alongside the idea of relative distinctiveness (e.g., Forrin et al., 2012; MacLeod et al., 2010; MacLeod & Bodner, 2017; Zhou & MacLeod, 2021). Importantly, however, the diagnostic value of this heuristic does not depend on distinctiveness relative to silent items; it merely requires successful recollection of the productive experience (Zhou & MacLeod, 2021).

present during encoding (indeed, one description of the distinctiveness heuristic states that participants mentally "replay" the encoded experience; MacLeod et al., 2010, p. 673). From this perspective, my findings from Chapter 2 and Chapter 4—which indicated reinstatement of encoding-related processes—are arguably compatible with the heuristic account.

More recent work (Wakeham-Lewis et al., 2022) has proposed that participants' recognition judgements may be guided by mental simulation: one may imagine reading a presented word aloud, and if the imagined experience matches encoded information associated with that word, then it may be endorsed as "old". What sort of neural predictions might we draw from this theory? One could argue that, because the simulated experience must eventually be matched against encoded information, this process should still entail some degree of reinstatement, similar to the recollective component of the heuristic account. Moreover, univariate fMRI work has revealed that imagining (i.e., mentally simulating) an action recruits many of the same cortical areas as observing or physically performing that action, notably including premotor and somatosensory cortices responsible for action planning and perceptual feedback (see Hardwick et al., 2018 for a meta-analysis). From this perspective, we might predict that simulated reading aloud (during retrieval) should elicit similar patterns of activation to those elicited during actual reading aloud (during encoding). As such, the evidence of reinstatement observed in Chapters 2 and 4 are also consistent with Wakeham-Lewis and colleagues' account.

5.4.2. Divergence From Prior Accounts

Two major findings from this dissertation cannot be easily accounted for by prior literature surrounding the production effect. First, my multivariate analyses (Chapters 3 & 4) revealed a number of processes that were more evident in the silent reading condition compared to aloud reading. Chapter 3 revealed greater decodability of orthographic information during silent reading, while Chapter 4 revealed that recognising silent words is associated with more extensive neural reactivation (both effects were

relative to the aloud condition). These findings are difficult to reconcile with the cognitive accounts described above, because a common theme across all of them is that reading aloud represents, in one form or another, a heightened cognitive state compared to reading silently: processing is either more distinctive (Forrin et al., 2012; Jamieson et al., 2016; MacLeod et al., 2010; MacLeod & Bodner, 2017), deeper (Fawcett et al., 2022), more attention-demanding (Fawcett & Ozubko, 2016; Mama et al., 2018; Mama & Icht, 2019; Ozubko, Gopie, et al., 2012), or otherwise *increased*[22]. This general view suggests that perceptual and memory-related processes should be unilaterally enhanced in the aloud reading condition. Stated differently: the traditional view holds that any cognitive process which is present for aloud reading should either be present to the same extent, diminished, or absent entirely for silent reading. My findings stand in contrast to this view because they reveal that aloud and silent reading engender dissociable neural states, with some processes (namely orthographic representation and neural reactivation) being more prominent in the silent condition. These findings provide a perspective which has not been expressed in prior literature. I will elaborate on the implications of this perspective in the next section (Section 5.4.3).

Secondly, Chapter 4 revealed that successful recognition of aloud words depends, to some extent, on neural transformation. Neural transformation entails systematic changes in activation patterns elicited by the same stimulus; in this context, the change occurs between encoding and retrieval. By contrast, accounts of the production effect either assume that encoded information is retrieved in its original format, or they do not specify one way or another. For example, MacLeod et al. (2010, p. 673) describe encoded speech information being "replayed" during retrieval (implying reinstatement of information its original format); while Fawcett et al. (2022) propose that semantic features may be diagnostic as to whether a word was previously studied, but do not

[22] The only exception, to my knowledge, is the item-order account described by Jonker et al. (2014), which describes enhanced relational processing during silent reading (see Section 3). However, this account is exclusively concerned with recall memory in pure-list designs, and therefore not relevant to the mixed-list production effect in recognition (the focus of this dissertation).

specify how these features are recalled from memory. Such descriptions are difficult to reconcile with the notion of encoded information being explicitly changed or reorganised. In Chapter 4 I made the case that the observed transformation effect may reflect reorganisation of retrieved speech-related information, perhaps reflecting the use of an evaluative heuristic (MacLeod & Bodner, 2017) or comparison between simulated and encoded information (Wakeham-Lewis et al., 2022). While the results are arguably compatible with strategy-based accounts (based on the above reasoning), my proposal that these processes might entail transformation of encoded information is entirely novel, because (to my knowledge) no prior work concerning the production effect has made this possibility explicit. As an aside, the notion that transformation is driven by one's internal goals (e.g., encoding versus mental evaluation) is entirely consistent with the wider literature surrounding memory retrieval (e.g., Favila et al., 2018); I will elaborate on this point in Section 5.5.2.

5.4.3. Towards A More Integrated Account of The Production Effect.

My findings present some interesting implications for future work on the production effect. First and foremost, I have shown that the multiple processes likely give rise to the recognition advantage for words that were read aloud. During encoding, there is evidence for roles of relative distinctiveness, attention, and semantic processing. During retrieval, it seems that both reactivation and transformation contribute to successful remembering. The notion that multiple processes might contribute to the production effect is not new, and has been expressed by multiple authors (e.g., Fawcett et al., 2022; Fawcett & Ozubko, 2016; Taikh & Bodner, 2016). However, this idea was previously based exclusively on behavioural data (that is, modulation of the recognition-based production effect by various experimental manipulations). The value of my work is that it offers the first evidence for multiple processes at the neural level, thus providing convergent evidence for this claim.

Second—and perhaps more controversially—I have shown that some neural (and, by implication, mental) processes are enhanced during silent reading relative to aloud reading. This finding should motivate a more nuanced perspective of cognitive processes surrounding the production effect. Rather than emphasising ways in which aloud reading is "special" (and, by extension, how silent reading is "not special"), I suggest that researchers simply consider any encoding task to engender different cognitive states depending on the reader's goals. Certainly, information that is necessary for (or elicited by) speech production appears to be useful for retrieval, and thus confers a recognition advantage to words that were read aloud. Equally, however, the cognitive state(s) engendered by reading silently may lead to other interesting consequences (e.g., Jonker et al., 2014 showed that item order information is better preserved following silent reading compared with aloud reading). As an aside, I do not consider the emphasis on aloud processing to be a shortcoming of prior work; virtually all of the theory surrounding the production effect has been based on behavioural data which consistently reveals a memory advantage for aloud words relative to silent. It is therefore only natural that researchers should emphasise processing benefits of aloud reading. My point is that neuroimaging may offer complementary insights to those of purely behavioural work. In this case neuroimaging revealed that aloud and silent reading entail dissociable, rather than one-sided encoding and retrieval processes.

Overall, I feel that future work may benefit from a more integrated view of cognition surrounding the production effect; one which considers that multiple processes likely contribute to behavioural memory outcomes, and which also considers the consequences of dynamic cognitive states engendered by these two tasks. One final comment I will make here is that neuroimaging clearly offers promise for better understanding the production effect. In some cases it may offer convergent support for existing claims derived from behavioural data (as discussed in Section 5.4.1); in others, it can provide novel insights into underlying cognitive processes (Section 5.4.2). As such, I feel that future work on the production effect will benefit from combined behavioural and neuroimaging investigation.

5.5. Convergence With Neurobiological Accounts of Episodic Memory

This dissertation is not only relevant to the production effect; my findings (particularly those from Chapters 2 & 4) also resonate with the broader literature on neural correlates of episodic encoding and retrieval.

5.5.1. Reactivation Accounts

Episodic memory has long been thought to entail "mental time travel" into the past (Tulving, 1985). Contemporary accounts have formalised this view by specifying that successful remembering should, at least to some extent, entail recapitulation of perceptual states that were present during encoding (e.g., Matheson & Barsalou, 2018; Meyer & Damasio, 2009; Xue, 2018). Much univariate fMRI work has met this prediction, revealing that retrieval recruits many of the same cortical areas as were activated during the initial encoding experience (see Danker & Anderson, 2010 for review). Such effects are often reflective of modality- or context-specific encoding conditions; for example, recalling an auditory stimulus or a picture will elicit activation in cortices that were differentially engaged when perceiving each respective modality (e.g., Wheeler et al., 2000). More recently, and as discussed in Chapter 1, multivariate studies have demonstrated that stimulus-specific activation patterns from encoding may be reinstated during retrieval (Favila et al., 2018; Jonker et al., 2018; Kuhl & Chun, 2014; Ritchey et al., 2013; Staresina et al., 2012; Tompary et al., 2016; Trelle et al., 2019; Wing et al., 2015; Xiao et al., 2017). Here, reactivation often reflects contextual processing demands that were present at encoding, but partially or entirely absent during retrieval. A good example comes from Xiao et al. (2017). Here, participants studied word-picture pairs during encoding; importantly, each picture was paired with multiple unique words. During retrieval, participants saw a cue (word) presented on screen and were instructed to recall

details of the associated image[23]. The authors computed correlations between pairs of encoding and retrieval trials and report that, in bilateral occipital cortex, correlations were significantly higher for pairs that shared the same picture but used a different cue (C^-P^+ pairs in [4]) compared with pairs that did not share cue or picture (C^-P^-). These findings indicate that image-specific activation patterns from encoding were reinstated during retrieval, despite there being no perceptual overlap between encoding and retrieval trials. Such findings from multivariate studies have provided further support for the notion that remembering entails mental reinstatement of earlier perceptual states (also see Xue, 2018 for review and discussion).

Some findings from Chapters 2 and 4 are entirely consistent with the above literature. In Chapter 2 I found that reading words aloud and later recognizing those words elicited overlapping activation in frontal and temporal cortices (which are broadly associated with speech production), consistent with prior univariate work. Importantly, this reactivation effect was greater for the aloud reading condition than the silent reading condition, mirroring the context-dependent effects seen by previous studies. Similarly, in Chapter 4 I showed that successfully remembering aloud words (compared with silent words) was associated with reactivation in the left insula, which has been linked to sensorimotor and auditory processing during speech production (Woolnough et al., 2019). Again, this is consistent with prior work showing that perceptual processes recruited during encoding are selectively reactivated at retrieval.

My reactivation analyses in Chapter 4 did produce one unexpected result: words in the silent reading condition were associated with more reactivation (relative to aloud reading) in ventral temporal areas associated with visual analysis and orthographic processing (including putative VWFA, which is widely regarded as being central to perceiving visual word-forms, as discussed in Chapter 3). That there should be any

[23] Thus, encoding and retrieval trials could be paired according to whether they shared the same cue and picture (C^+P^+ pairs), different cue but same picture (C^-P^+), or different cue and different picture (C^-P^-).

between-condition reactivation differences in these areas is surprising; based on the literature described above, one would expect any differences to be driven by non-overlapping processing demands during encoding. Because visual and orthographic demands are shared by both conditions (participants must always read the word, regardless of whether or not they need to say it aloud), this result is hard to explain in terms of context-specific encoding. In Chapter 4 I suggested that participants may utilise visual / orthographic information when making recognition decisions on silent words because relatively less information is available to inform that decision. Stated differently, in the absence of speech-related information afforded to aloud items, participants must rely more on what limited information *was* encoded. As for implications on the wider field, this finding suggests that reactivation is not only subject to context-specific processing during encoding, but is also modulated by the relative usefulness of encoded information (but see Section 5.2). When encoding entails multisensory processing (as in aloud reading), retrieval dynamics favour the most diagnostic information (consistent with distinctiveness-based accounts of the production effect, discussed earlier). Alternatively, when encoded information is relatively sparse (as in silent reading), participants must make do with what is available.

5.5.2. Transformation Accounts

Another view of memory retrieval holds that remembering entails systematic changes in the way that encoded information is represented at the neural level (Favila et al., 2020; Xue, 2022). This view is not incompatible with reactivation accounts; for example Favila et al. (2020) emphasise that reactivation and transformation are both important components of memory retrieval.

In the context of fMRI, transformation effects may manifest as a shift in the spatial location at which stimulus / item information can be decoded (Favila et al., 2018; Long & Kuhl, 2021; Xiao et al., 2017). Alternatively, transformation may be captured within a patch of cortex by diminished pattern similarity between encoding and retrieval,

compared with that of repeated presentations *within* encoding or retrieval phases (J. Chen et al., 2017; Xiao et al., 2017); in Chapter 4, I captured transformation using this latter approach. Transformation by either measure is generally thought to reflect systematic reorganisation of encoded information, perhaps as an intrinsic component of effective encoding (Liu et al., 2021), or to align with specific retrieval goals. With respect to the latter possibility, a good example comes from Favila et al. (Favila et al., 2018; Experiment 2). These authors reported that stimulus features (colour or taxonomic class), initially decodable in occipitotemporal cortex during encoding, were decodable in lateral parietal cortex during retrieval (i.e., between-region transformation). Critically, feature decodability during retrieval depended on which feature was relevant to the current task (e.g., stimulus colour was more decodable when participants were instructed to remember colour information)—hence, transformation was modulated by participants' goals. This finding is consistent with the more general observation that feature decodability often aligns with participants' current task demands (Kuhl et al., 2013; Long & Kuhl, 2021; Nastase et al., 2017).

My findings from Chapter 4—which revealed differential degrees of encoding-retrieval transformation between the aloud and silent reading conditions—provide an interesting contrast to the research described above. Previous work has shown that transformation is driven by the explicit retrieval task, whereas in my experiment, transformation was modulated by *preceding* encoding conditions. In that chapter and above (Section 5.1) I argued that participants' judgements on the recognition task may depend on different sources of information; one could argue that the observed transformation effect was driven by intrinsic cognitive states (i.e., recollecting speech information versus visual / orthographic information) which arose due to the preceding encoding conditions.

To my knowledge, only one other study has revealed similar effects of intrinsic states (as opposed to explicit task demands) on neural transformation. Long and Kuhl (2021) presented participants with images of objects and, on each trial, instructed them

to either commit the image to memory (encode instruction) or to recall a different image of the same object that they had viewed prior to scanning (retrieve instruction). These authors were able to characterise intrinsic cognitive states across large-scale attentional networks, either biased towards encoding or retrieval, on each trial. Importantly, these intrinsic states were dissociable from the actual instruction presented to participants[24]. These authors reported that during encoding states, item identity (that is, the correlation between patterns elicited by presentations of the same stimulus minus those of different stimuli) was more decodable in ventral temporal cortex compared with lateral parietal cortex; during retrieval states, however, this effect was reversed. These results indicate that retrieval-related transformation (that is, a shift in the location of content representations, from ventral temporal to parietal cortices) may be driven not only by explicit task demands, but also by covert cognitive states. I take a similar view of my findings from Chapter 4—in this case, I suggest that recognizing aloud and silent words may entail rather different cognitive states, giving rise to different degrees of transformation. More generally, Chapter 4 provides support for the claim that task-dependent changes in neural representation need not depend on explicit task demands during retrieval.

5.5.3. Tradeoffs Between Reactivation and Transformation

To my knowledge, no prior work has explicitly addressed potential tradeoffs between reactivation and transformation. In their review of transformation effects in

[24] These intrinsic states were derived from the output of a classifier that was applied to decode the instruction on each trial (encode or retrieve) from the elicited fMRI patterns in each network. While the classifier made a binary decision for each trial (labelling it as "encode" or "retrieve"), it also quantified *classifier evidence* for each trial; that is, a continuous measure reflecting the probability that the classifier assigned the correct label on that trial (this measure has been conceptualised as the classifier's confidence in its decision; D. E. Smith et al., 2022). Long and Kuhl (2021) used this measure to infer intrinsic cognitive states: neural responses to items were median-split according to their classifier evidence, separately within each instructional condition. Trials with high classifier evidence (above the median) had a high probability of being labelled as "encode", and therefore the elicited network-wide pattern was considered to be more encoding-like; conversely, low classifier evidence (below the median) was taken as an indicator of a retrieval-like state.

neuroimaging literature, Favila and colleagues make this issue explicit by posing a number of outstanding questions (I have numbered each question so that I may address them in turn below):

> "[1] What determines the relative degree of neural reactivation versus transformation across brain regions observed during memory retrieval? … [2] Does the relative degree of reactivation versus transformation depend on whether memory tasks involve recall versus recognition judgments? [3] Do reactivation and transformation trade-off or are they independent?" (Favila et al., 2020, p. 9).

While these authors were explicitly concerned with transformation across brain regions, their questions are still applicable to within-region transformation. My findings offer a starting point for answering some of these questions. Most concretely, they indicate that reactivation and transformation are not completely independent [3], since I observed differential contributions of each during successful recognition, depending on the preceding encoding condition. With respect to [2], this is difficult to answer based on my data, since my experiment entailed a recognition task only. One could make the case that, since aloud trials are thought to elicit recollection of the encoded speech experience (MacLeod & Bodner, 2017), recognising aloud words entails similar retrieval dynamics to a recall task. However, [2] could be much more easily tested by explicitly comparing processes across recognition and recall tasks.

As for *what* drives the tradeoff [1], again this likely has something to do with internal retrieval processes driving recognition in each condition. If, as argued in Chapter 4 and above, recollection and evaluation of speech information underlies transformation on aloud trials, and if reinstatement of visual / orthographic information underlies reactivation on silent trials, then ultimately the tradeoff between reactivation and

transformation must be driven by how much information is available and useful in the encoded memory trace (at least in the context of this study). Stated differently, the tradeoff may be modulated by distinctiveness: aloud words are associated with rich speech information, which may be recollected and evaluated to endorse those words as "old". By contrast, encoded information for silent words is relatively sparse, forcing participants to depend on mental reinstatement of visual / orthographic information. Admittedly, my answer to [1] is largely conjectured, and depends on my own assumptions about what retrieval processes are at play during recognition of aloud and silent words. As such, further work is needed to elucidate the nature of this tradeoff.

5.6. Future Directions

Future research may build upon this dissertation in a number of ways. First and foremost, I strongly encourage future neuroimaging-based investigations of the production effect to make use of multivariate analyses. As discussed in Chapter 1, these techniques are informative as to the informational content of neural responses, and so offer detailed insight into processes which contribute to encoding and retrieval of individual stimuli.

Future research may build on the findings of Chapter 3 by linking decodability (of specific stimulus properties) to behavioural memory outcomes. As I noted in that chapter and in Section 4.2, the observed task-dependent changes in decodability may be epiphenomenal to the production effect. As such, an explicit link to behavioural memory performance is needed to verify the mnemonic relevance of articulatory, semantic, etc. information. For example, investigating whether decodability of articulatory information is positively correlated with recognition performance would be directly informative as to the role of encoded speech information in the production effect. Such an investigation would be relatively straightforward and could use a similar paradigm to that of Chapter 3; the only caveat being that it would require a sufficiently large stimulus list to detect the production effect (that is, the stimulus list would need to be large enough to eliminate

ceiling effects on memory performance, given the need to repeat each stimulus multiple times; see Section 3).

Chapter 4 highlighted the possibility that different encoding conditions may give rise to a tradeoff between reactivation and transformation during retrieval. Further work is needed to clarify this issue. For example, to test the possibility that this tradeoff is modulated by distinctiveness (as I suggested in the preceding section), future research might entail a paradigm which systematically manipulates the degree of distinctiveness conferred upon study items (e.g., singing aloud, reading aloud, mouthing silently, as in Forrin et al., 2012; Quinlan & Taylor, 2013, 2019). Moreover, given the reliable presence of both reactivation and transformation when stimuli are exclusively in the visual modality (e.g., Favila et al., 2020; Liu et al., 2021; Xiao et al., 2017), tradeoffs should be observable outside the context of a production paradigm—for example, by manipulating the complexity (and therefore distinctiveness) of encoded visual scenes.

Another potential avenue would be to explore between-region transformation during retrieval. The whole-brain searchlight approach which I adopted in Chapter 4 is ill-suited to such an analysis; however, between-region transformation is relatively straightforward to assess using a-priori ROIs (Favila et al., 2018; Long & Kuhl, 2021; Xiao et al., 2017). In Chapter 4 I suggested that encoded speech information—initially present in the insula (given that this area exhibited reactivation effects) or possibly medial frontal cortices (given that articulatory information was decodable in these areas in Chapter 3)—might be reinstated in the precuneus during retrieval. This possibility might be directly tested with RSA: one could compute representational dissimilarity matrices (RDMs) from activity patterns in the insula or medial frontal cortices (along with any other ROIs which seem appropriate) during encoding, and then correlate these with RDMs derived from the precuneus during retrieval. Evidence of a positive correlation between RDMs from these speech-related regions (during encoding) and the precuneus (during retrieval) would provide supporting evidence for the above claim.

It may also be worth applying MVPA to retrieval data acquired from a remember-know recognition paradigm (Tulving, 1985)[25]. This paradigm differentiates recollection (remembering a specific event, with accompanying contextual information) from familiarity (a general sense of "knowing" that one has experienced the stimulus, without explicit recollection)—indeed, this distinction has received much attention in the production effect literature (e.g., Fawcett & Ozubko, 2016; Ozubko, Gopie, et al., 2012). For example, it may be the case that reactivation and transformation contribute differentially to familiarity and recollection—this would provide some valuable context for my own results, since old / new judgments may be driven by a combination of these two processes. Such a finding would also be informative as to the broader roles of reactivation and transformation in episodic memory.

5.7. Final Remarks

The production effect is, at face value, a simple phenomenon: words read aloud are better remembered than words read silently. Apparent neural substrates of this effect, however, are far from straightforward. In this dissertation I have presented evidence that, compared to reading silently, reading aloud entails engagement of speech-related sensorimotor processes, along with enhanced semantic processing and (possibly) greater attentional engagement. All of these processes likely confer a benefit to encoding. Furthermore, I have shown that retrieval of words previously read aloud entails both neural reactivation (indicative of mental reinstatement of speech processes) and transformation (perhaps reflecting evaluation of the reinstated content). It is likely that the recognition advantage for aloud words arises due to a combination of all the aforementioned mental processes. Future neuroimaging research will hopefully elaborate on these findings. Moreover, I hope that such work will help to guide formal cognitive accounts of the production effect. By developing cognitive theory that is

[25] Although this paradigm was used in the experiment from Chapter 2, the parameters of that experiment were not well suited to multivariate analysis (Zeithamova et al., 2017).

grounded in observed neural phenomena (as measured by fMRI or otherwise), we will hopefully arrive at a more integrated account of the production effect; one which respects the roles of multiple cognitive processes, and which is also capable of explaining all apparent neural correlates of the production effect.

6. BIBLIOGRAPHY

Alario, F.-X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research*, *1076*(1), 129–143. https://doi.org/10.1016/j.brainres.2005.11.104

Alberton, B. A. V., Nichols, T. E., Gamba, H. R., & Winkler, A. M. (2020). Multiple testing correction over contrasts for brain imaging. *NeuroImage*, *216*, 116760. https://doi.org/10.1016/j.neuroimage.2020.116760

Arichi, T., Fagiolo, G., Varela, M., Melendez-Calderon, A., Allievi, A., Merchant, N., Tusor, N., Counsell, S. J., Burdet, E., & Beckmann, C. F. (2012). Development of BOLD signal hemodynamic responses in the human brain. *Neuroimage*, *63*(2), 663–673.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Bailey, L. M., Bodner, G. E., Matheson, H. E., Stewart, B. M., Roddick, K., O'Neil, K., Simmons, M., Lambert, A. M., Krigolson, O. E., Newman, A. J., & Fawcett, J. M. (2021). Neural correlates of the production effect: An fMRI study. *Brain and Cognition*, *152*, 105757. https://doi.org/10.1016/j.bandc.2021.105757

Bailey, L. M., Lockary, K., & Higby, E. (2023). Cross-linguistic influence in the bilingual lexicon: Evidence for ubiquitous facilitation and context-dependent interference effects on lexical processing. *Bilingualism: Language and Cognition*, 1–20. https://doi.org/10.1017/S1366728923000597

Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., & Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences*, *104*(21), 9087–9092. https://doi.org/10.1073/pnas.0703300104

Beijering, K., Gooskens, C., & Heeringa, W. (2008). Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm. *Linguistics in the Netherlands*, *25*(1), 13–24.

Binder, J. R., & Desai, R. H. (2011). The Neurobiology of Semantic Memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bolger, D. J., Hornickel, J., Cone, N. E., Burman, D. D., & Booth, J. R. (2008). Neural correlates of orthographic and phonological consistency effects in children. *Human Brain Mapping*, *29*(12), 1416–1429. https://doi.org/10.1002/hbm.20476

Bone, M. B., & Buchsbaum, B. R. (2021). Detailed episodic memory depends on concurrent reactivation of basic visual features within the posterior hippocampus and early visual cortex. *Cerebral Cortex Communications*, *2*(3), tgab045.

Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: A perceptual to conceptual gradient of semantic coding. *NeuroImage*, *143*, 128–140. https://doi.org/10.1016/j.neuroimage.2016.08.068

Bosch, S. E., Jehee, J. F. M., Fernández, G., & Doeller, C. F. (2014). Reinstatement of Associative Memories in Early Visual Cortex Is Signaled by the Hippocampus. *Journal of Neuroscience*, *34*(22), 7493–7500. https://doi.org/10.1523/JNEUROSCI.0805-14.2014

Bourguignon, N. J. (2014). A rostro-caudal axis for language in the frontal lobe: The role of executive control in speech production. *Neuroscience & Biobehavioral Reviews*, *47*, 431–444. https://doi.org/10.1016/j.neubiorev.2014.09.008

Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, *107*(17), 7939–7944. https://doi.org/10.1073/pnas.0904402107

Carota, F., Nili, H., Pulvermüller, F., & Kriegeskorte, N. (2021). Distinct fronto-temporal substrates of distributional and taxonomic similarity among words: Evidence from RSA of BOLD signals. *NeuroImage*, *224*, 117408. https://doi.org/10.1016/j.neuroimage.2020.117408

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, *7*(3), Article 3. https://doi.org/10.1038/s41562-022-01516-2

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583. https://doi.org/10.1093/brain/awl004

Chee, Q. W., Chow, K. J., Yap, M. J., & Goh, W. D. (2020). Consistency norms for 37,677 english words. *Behavior Research Methods*, *52*(6), 2535–2555. https://doi.org/10.3758/s13428-020-01391-7

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), Article 1. https://doi.org/10.1038/nn.4450

Chen, L., Wassermann, D., Abrams, D. A., Kochalka, J., Gallardo-Diez, G., & Menon, V. (2019). The visual word form area (VWFA) is part of both language and attention circuitry. *Nature Communications*, *10*, 5601. https://doi.org/10.1038/s41467-019-13634-z

Ciaramelli, E., Grady, C. L., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: A hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*(7), 1828–1851. https://doi.org/10.1016/j.neuropsychologia.2008.03.022

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, *125*(5), 1054–1069. https://doi.org/10.1093/brain/awf094

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cornelissen, P. L., Kringelbach, M. L., Ellis, A. W., Whitney, C., Holliday, I. E., & Hansen, P. C. (2009). Activation of the Left Inferior Frontal Gyrus in the First 200 ms of Reading: Evidence from Magnetoencephalography (MEG). *PLoS ONE*, *4*(4), e5359. https://doi.org/10.1371/journal.pone.0005359

Cutini, S., Scatturin, P., Menon, E., Bisiacchi, P. S., Gamberini, L., Zorzi, M., & Dell'Acqua, R. (2008). Selective activation of the superior frontal gyrus in task-switching: An event-related fNIRS study. *NeuroImage*, *42*(2), 945–955. https://doi.org/10.1016/j.neuroimage.2008.05.013

Dadario, N. B., & Sughrue, M. E. (2023). The functional role of the precuneus. *Brain*, *146*(9), 3598–3607. https://doi.org/10.1093/brain/awad181

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, *136*(1), 87.

Danker, J. F., Tompary, A., & Davachi, L. (2017). Trial-by-Trial Hippocampal Encoding Activation Predicts the Fidelity of Cortical Reinstatement During Subsequent Retrieval. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(7), 3515–3524. https://doi.org/10.1093/cercor/bhw146

Davis, T., Xue, G., Love, B. C., Preston, A. R., & Poldrack, R. A. (2014). Global Neural Pattern Similarity as a Common Basis for Categorization and Recognition Memory. *Journal of Neuroscience*, *34*(22), 7472–7484. https://doi.org/10.1523/JNEUROSCI.3376-13.2014

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational Similarity Analysis Reveals Commonalities and Differences in the Semantic Processing of Words and Objects. *Journal of Neuroscience*, *33*(48), 18906–18916. https://doi.org/10.1523/JNEUROSCI.3809-13.2013

Diedrichsen, J., Ridgway, G. R., Friston, K. J., & Wiestler, T. (2011). Comparing the similarity and spatial structure of neural representations: A pattern-component model. *NeuroImage*, *55*(4), 1665–1678. https://doi.org/10.1016/j.neuroimage.2011.01.044

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89.

Dietz, N. A. E., Jones, K. M., Gareau, L., Zeffiro, T. A., & Eden, G. F. (2005). Phonological decoding involves left posterior fusiform gyrus. *Human Brain Mapping*, *26*(2), 81–93. https://doi.org/10.1002/hbm.20122

Dinteren, R. van, Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 Development across the Lifespan: A Systematic Review and Meta-Analysis. *PLOS ONE*, *9*(2), e87347. https://doi.org/10.1371/journal.pone.0087347

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(1), 224–240. https://doi.org/10.1037//0278-7393.21.1.224

Fandakova, Y., Johnson, E. G., & Ghetti, S. (2021). Distinct neural mechanisms underlie subjective and objective recollection and guide memory-based decision making. *eLife*, *10*, e62520. https://doi.org/10.7554/eLife.62520

Favila, S. E., Lee, H., & Kuhl, B. A. (2020). Transforming the Concept of Memory Reactivation. *Trends in Neurosciences*, *43*(12), 939–950. https://doi.org/10.1016/j.tins.2020.09.006

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal Representations of Stimulus Features Are Amplified during Memory Retrieval and Flexibly Aligned with Top-Down Goals. *The Journal of Neuroscience*, *38*(36), 7809–7821. https://doi.org/10.1523/JNEUROSCI.0564-18.2018

Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*(1), 1–5. https://doi.org/10.1016/j.actpsy.2012.10.001

Fawcett, J. M., Baldwin, M. M., Whitridge, J. W., Swab, M., Malayang, K., Hiscock, B., Drakes, D. H., & Willoughby, H. V. (2023). Production improves recognition and reduces intrusions in between-subject designs: An updated meta-analysis. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *77*(1), 35–44. https://doi.org/10.1037/cep0000302

Fawcett, J. M., Bodner, G. E., Paulewicz, B., Rose, J., & Wakeham-Lewis, R. (2022). Production can enhance semantic encoding: Evidence from forced-choice recognition with homophone versus synonym lures. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-022-02140-x

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 99.

Fawcett, J. M., Quinlan, C. K., & Taylor, T. L. (2012). Interplay of the production and picture superiority effects: A signal detection analysis. *Memory*, *20*(7), 655–666.

Fiez, J. A., Balota, D. A., Raichle, M. E., & Petersen, S. E. (1999). Effects of Lexicality, Frequency, and Spelling-to-Sound Consistency on the Functional Anatomy of Reading. *Neuron*, *24*(1), 205–218. https://doi.org/10.1016/S0896-6273(00)80833-8

Fiez, J. A., & Petersen, S. E. (1998). Neuroimaging studies of word reading. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 914–921. https://doi.org/10.1073/pnas.95.3.914

Fine, J. M., & Hayden, B. Y. (2021). The whole prefrontal cortex is premotor cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1844), 20200524. https://doi.org/10.1098/rstb.2020.0524

Fischer-Baum, S., Bruggemann, D., Gallego, I. F., Li, D. S. P., & Tamez, E. R. (2017). Decoding levels of representation in reading: A representational similarity approach. *Cortex*, *90*, 88–102. https://doi.org/10.1016/j.cortex.2017.02.017

Fischer-Baum, S., Kook, J. H., Lee, Y., Ramos-Nuñez, A., & Vannucci, M. (2018). Individual Differences in the Neural and Cognitive Mechanisms of Single Word Reading. *Frontiers in Human Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnhum.2018.00271

Fontan, L., Ferrané, I., Farinas, J., Pinquier, J., & Aumont, X. (2016). Using phonologically weighted Levenshtein distances for the prediction of microscopic intelligibility. *Annual Conference Interspeech (INTERSPEECH 2016)*, 650–654.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*(7), 1046–1055. https://doi.org/10.3758/s13421-012-0210-8

Freund, M. C., Bugg, J. M., & Braver, T. S. (2021). A Representational Similarity Analysis of Cognitive Control during Color-Word Stroop. *Journal of Neuroscience*, *41*(35), 7388–7402. https://doi.org/10.1523/JNEUROSCI.2956-20.2021

Garrido, L., Vaziri-Pashkam, M., Nakayama, K., & Wilmer, J. (2013). The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Frontiers in Neuroscience*, *7*. https://www.frontiersin.org/articles/10.3389/fnins.2013.00174

Genon, S., Li, H., Fan, L., Müller, V. I., Cieslik, E. C., Hoffstaedter, F., Reid, A. T., Langner, R., Grefkes, C., Fox, P. T., Moebus, S., Caspers, S., Amunts, K., Jiang, T., & Eickhoff, S. B. (2017). The Right Dorsal Premotor Mosaic: Organization, Functions, and Connectivity. *Cerebral Cortex (New York, NY)*, *27*(3), 2095–2110. https://doi.org/10.1093/cercor/bhw065

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Gramfort, A., Pallier, C., Varoquaux, G., & Thirion, B. (2012). Decoding Visual Percepts Induced by Word Reading with fMRI. *2012 Second International Workshop on Pattern Recognition in NeuroImaging*, 13–16. https://doi.org/10.1109/PRNI.2012.20

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), Article 8. https://doi.org/10.1038/nrn3747

Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*, *188*, 668–679. https://doi.org/10.1016/j.neuroimage.2018.12.046

Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2019). Untangling featural and conceptual object representations. *NeuroImage*, *202*, 116083. https://doi.org/10.1016/j.neuroimage.2019.116083

Guo, W., Geng, S., Cao, M., & Feng, J. (2022). Functional Gradient of the Fusiform Cortex for Chinese Character Recognition. *eNeuro*, *9*(3). https://doi.org/10.1523/ENEURO.0495-21.2022

Hall, K. C., Mackie, J. S., & Lo, R. Y.-H. (2019). Phonological CorpusTools: Software for doing phonological analysis on transcribed corpora. *International Journal of Corpus Linguistics*, *24*(4), 522–535. https://doi.org/10.1075/ijcl.18009.hal

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience & Biobehavioral Reviews*, *94*, 31–44.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review*, *111*, 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Hasinski, A. E., & Sederberg, P. B. (2016). Trial-level information for individual faces in the fusiform face area depends on subsequent memory. *NeuroImage*, *124*(Pt A), 526–535. https://doi.org/10.1016/j.neuroimage.2015.08.065

Hassall, C. D., Quinlan, C. K., Turk, D. J., Taylor, T. L., & Krigolson, O. E. (2016). A preliminary investigation into the neural basis of the production effect. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 139.

Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word reading: Evidence from fast behavioral and brain responses. *NeuroImage*, *60*(2), 1462–1477. https://doi.org/10.1016/j.neuroimage.2012.01.061

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Hayes, B. (2008). *Introductory phonology* (Vol. 7). John Wiley & Sons.

Henningsen-Schomers, M. R., & Pulvermüller, F. (2022). Modelling concrete and abstract concepts using brain-constrained deep neural networks. *Psychological Research*, *86*(8), 2533–2559. https://doi.org/10.1007/s00426-021-01591-6

Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *19*(10), 3962–3972. https://doi.org/10.1523/JNEUROSCI.19-10-03962.1999

Hertrich, I., Dietrich, S., & Ackermann, H. (2016). The role of the supplementary motor area for speech and language processing. *Neuroscience & Biobehavioral Reviews*, *68*, 602–610. https://doi.org/10.1016/j.neubiorev.2016.06.030

Hertrich, I., Dietrich, S., Blum, C., & Ackermann, H. (2021). The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.645209

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), Article 5. https://doi.org/10.1038/nrn2113

Hintzman, D. L. (1976). Repetition and memory. *Psychology of Learning and Motivation*, *10*, 47–91.

Jafari, M., & Ansari-Pour, N. (2019). Why, When and How to Adjust Your P Values? *Cell Journal (Yakhteh)*, *20*(4), 604–607. https://doi.org/10.22074/cellj.2019.5992

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 154.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/s1053-8119(02)91132-8

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/s1361-8415(01)00036-6

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *Neuroimage*, *20*(2), 693–712.

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences*, *115*(43), 11084–11089. https://doi.org/10.1073/pnas.1800006115

Jonker, T. R., Levene, M., & MacLeod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 441.

Kaiser, D., Moeskops, M. M., & Cichy, R. M. (2018). Typical retinotopic locations impact the time course of object coding. *NeuroImage*, *176*, 372–379. https://doi.org/10.1016/j.neuroimage.2018.05.006

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Karis, D., Fabiani, M., & Donchin, E. (1984). “P300” and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, *16*(2), 177–216.

Khalilian-Gourtani, A., Wang, R., Chen, X., Yu, L., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., Wang, Y., & Flinker, A. (2022). A Corollary Discharge Circuit in Human Speech. *bioRxiv*, 2022.09.12.507590. https://doi.org/10.1101/2022.09.12.507590

Kim, C., Chung, C., & Kim, J. (2013). Task-dependent response conflict monitoring and cognitive control in anterior cingulate and dorsolateral prefrontal cortices. *Brain Research*, *1537*, 216–223. https://doi.org/10.1016/j.brainres.2013.08.055

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, *54*(3), 2446–2461. https://doi.org/10.1016/j.neuroimage.2010.09.045

Koechlin, E. (2011). Frontal pole function: What is specifically human? *Trends in Cognitive Sciences*, *15*(6), 241.

Konishi, S., Jimura, K., Asari, T., & Miyashita, Y. (2003). Transient Activation of Superior Prefrontal Cortex during Inhibition of Cognitive Set. *Journal of Neuroscience*, *23*(21), 7776–7782. https://doi.org/10.1523/JNEUROSCI.23-21-07776.2003

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Krekelberg, B. (2022). *BayesFactor* (v2.2.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7006300

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://www.frontiersin.org/articles/10.3389/neuro.06.004.2008

Kuhl, B. A., & Chun, M. M. (2014). Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. *Journal of Neuroscience*, *34*(23), 8051–8060. https://doi.org/10.1523/JNEUROSCI.4328-13.2014

Kuhl, B. A., Johnson, M. K., & Chun, M. M. (2013). Dissociable Neural Mechanisms for Goal-Directed Versus Incidental Memory Reactivation. *Journal of Neuroscience*, *33*(41), 16099–16109. https://doi.org/10.1523/JNEUROSCI.0207-13.2013

Lee, C.-Y., Tsai, J.-L., Kuo, W.-J., Yeh, T.-C., Wu, Y.-T., Ho, L.-T., Hung, D. L., Tzeng, O. J., & Hsieh, J.-C. (2004). Neuronal correlates of consistency and frequency effects on Chinese character naming: An event-related fMRI study. *Neuroimage*, *23*(4), 1235–1245.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Lehman, V. T., Black, D. F., Bernstein, M. A., & Welker, K. M. (2016). Temporal lobe anatomy: Eight imaging signs to facilitate interpretation of MRI. *Surgical and Radiologic Anatomy*, *38*(4), 433–443. https://doi.org/10.1007/s00276-015-1582-9

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin*, *140*(6), 1534–1555. https://doi.org/10.1037/a0037830

Levy, J., Pernet, C., Treserras, S., Boulanouar, K., Aubry, F., Démonet, J.-F., & Celsis, P. (2009). Testing for the Dual-Route Cascade Reading Model in the Brain: An fMRI Effective Connectivity Account of an Efficient Reading Style. *PLOS ONE*, *4*(8), e6675. https://doi.org/10.1371/journal.pone.0006675

Li, A., Yang, R., Qu, J., Dong, J., Gu, L., & Mei, L. (2022). Neural representation of phonological information during Chinese character reading. *Human Brain Mapping*, *43*(13), 4013–4029. https://doi.org/10.1002/hbm.25900

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-018-08080-2

Liu, J., Zhang, H., Yu, T., Ren, L., Ni, D., Yang, Q., Lu, B., Zhang, L., Axmacher, N., & Xue, G. (2021). Transformative neural representations support long-term episodic memory. *Science Advances*, *7*(41), eabg9715. https://doi.org/10.1126/sciadv.abg9715

Liuzzi, A. G., Aglinskas, A., & Fairhall, S. L. (2020). General and feature-based semantic representations in the semantic network. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-65906-0

Loh, K. K., Procyk, E., Neveu, R., Lamberton, F., Hopkins, W. D., Petrides, M., & Amiez, C. (2020). Cognitive control of orofacial motor and vocal responses in the ventrolateral and dorsomedial human frontal cortex. *Proceedings of the National Academy of Sciences*, *117*(9), 4994–5005. https://doi.org/10.1073/pnas.1916459117

Long, N. M., & Kuhl, B. A. (2021). Cortical Representations of Visual Stimuli Shift Locations with Changes in Memory States. *Current Biology*, *31*(5), 1119-1126.e5. https://doi.org/10.1016/j.cub.2021.01.004

Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *NeuroImage*, *27*(4), 824–834. https://doi.org/10.1016/j.neuroimage.2005.05.008

Lundstrom, B. N., Petersson, K. M., Andersson, J., Johansson, M., Fransson, P., & Ingvar, M. (2003). Isolating the retrieval of imagined pictures during episodic memory: Activation of the left precuneus and left prefrontal cortex. *NeuroImage*, *20*(4), 1934–1943. https://doi.org/10.1016/j.neuroimage.2003.07.017

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*(5472), 1835–1838.

MacDonald, P. A., & MacLeod, C. M. (1998). The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica*, *98*(2–3), 291–310. https://doi.org/10.1016/s0001-6918(97)00047-4

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science*, *26*(4), 390–395.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Malonek, D., & Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. *Science*, *272*(5261), 551–554.

Mama, Y., Fostick, L., & Icht, M. (2018). The impact of different background noises on the Production Effect. *Acta Psychologica*, *185*, 235–242. https://doi.org/10.1016/j.actpsy.2018.03.002

Mama, Y., & Icht, M. (2019). Production Effect in Adults With ADHD With and Without Methylphenidate (MPH): Vocalization Improves Verbal Learning. *Journal of the International Neuropsychological Society: JINS*, *25*(2), 230–235. https://doi.org/10.1017/S1355617718001017

Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping*, *36*(5), 1963–1981. https://doi.org/10.1002/hbm.22749

Matheson, H. E., & Barsalou, L. W. (2018). Embodiment and Grounding in Cognitive Neuroscience. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–27). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119170174.epcn310

Matheson, H. E., Kenett, Y. N., Gerver, C., & Beaty, R. E. (2023). Representing creative thought: A representational similarity analysis of creative idea generation and evaluation. *Neuropsychologia*, *187*, 108587. https://doi.org/10.1016/j.neuropsychologia.2023.108587

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299. https://doi.org/10.1016/S1364-6613(03)00134-7

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual Perception. *The Journal of Neuroscience*, *33*(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. A., Patterson, K., McClelland, J. L., & Price, C. J. (2005). Dissociating reading processes on the basis of neuronal interactions. *Journal of Cognitive Neuroscience*, *17*(11), 1753–1765. https://doi.org/10.1162/089892905774589190

Meersmans, K., Storms, G., De Deyne, S., Bruffaerts, R., Dupont, P., & Vandenberghe, R. (2022). Orienting to different dimensions of word meaning alters the representation of word meaning in early processing regions. *Cerebral Cortex*, *32*(15), 3302–3317. https://doi.org/10.1093/cercor/bhab416

Mesulam, M.-M. (1998). From sensation to cognition. *Brain: A Journal of Neurology*, *121*(6), 1013–1052.

Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Neurosciences*, *32*(7), 376–382. https://doi.org/10.1016/j.tins.2009.04.002

Moerel, D., Grootswagers, T., Robinson, A. K., Shatek, S. M., Woolgar, A., Carlson, T. A., & Rich, A. N. (2022). The time-course of feature-based attention effects dissociated from temporal expectation and target-related processes. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-10687-x

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *Journal of Neuroscience*, *38*(14), 3534–3546. https://doi.org/10.1523/JNEUROSCI.2360-17.2018

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2022). *BayesFactor: Computation of Bayes Factors for Common Designs* (0.9.12-4.4) [Computer software]. https://cran.r-project.org/web/packages/BayesFactor/index.html

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, *103*, 130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026

Murphy, K. A., Jogia, J., & Talcott, J. B. (2019). On the neural basis of word reading: A meta-analysis of fMRI evidence using activation likelihood estimation. *Journal of Neurolinguistics*, *49*, 71–83. https://doi.org/10.1016/j.jneuroling.2018.08.005

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M. I., & Haxby, J. V. (2017). Attention Selectively Reshapes the Geometry of Distributed Semantic Representation. *Cerebral Cortex*, *27*(8), 4277–4291. https://doi.org/10.1093/cercor/bhx138

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective & Behavioral Neuroscience*, *12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Ogawa, S., Lee, T.-M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–9872.

Oh, A., Duerden, E. G., & Pang, E. W. (2014). The role of the insula in speech and language processing. *Brain and Language*, *135*, 96–103. https://doi.org/10.1016/j.bandl.2014.06.003

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*, 27.

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory (Hove, England)*, *20*(7), 717–727. https://doi.org/10.1080/09658211.2012.699070

Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*, *22*(5), 509–524.

Pasalar, S., Ro, T., & Beauchamp, M. S. (2010). TMS of posterior parietal cortex disrupts visual tactile multisensory integration. *European Journal of Neuroscience*, *31*(10), 1783–1790.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), Article 12. https://doi.org/10.1038/nrn2277

Peelen, M., & Downing, P. (2022). *Testing cognitive theories using multivariate pattern analysis of neuroimaging dat*. https://doi.org/10.31234/osf.io/rhzt9

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*, 177–183. https://doi.org/10.1016/j.neuropsychologia.2017.03.033

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*(2), 106–151. https://doi.org/10.1016/j.cogpsych.2010.04.001

Petersen, S. E., Fox, P. T., Snyder, A. Z., & Raichle, M. E. (1990). Activation of Extrastriate and Frontal Cortical Areas by Visual Words and Word-Like Stimuli. *Science*, *249*(4972), 1041–1044. https://doi.org/10.1126/science.2396097

Pflugshaupt, T., Gutbrod, K., Wurtz, P., von Wartburg, R., Nyffeler, T., de Haan, B., Karnath, H.-O., & Mueri, R. M. (2009). About the role of visual field defects in pure alexia. *Brain: A Journal of Neurology*, *132*(Pt 7), 1907–1917. https://doi.org/10.1093/brain/awp141

Pleisch, G., Karipidis, I. I., Brem, A., Röthlisberger, M., Roth, A., Brandeis, D., Walitza, S., & Brem, S. (2019). Simultaneous EEG and fMRI reveals stronger sensitivity to orthographic strings in the left occipito-temporal cortex of typical versus poor beginning readers. *Developmental Cognitive Neuroscience*, *40*, 100717. https://doi.org/10.1016/j.dcn.2019.100717

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia*, *48*(5), 1336–1342. https://doi.org/10.1016/j.neuropsychologia.2009.12.036

Polk, T. A., & Farah, M. J. (2002). Functional MRI evidence for an abstract, not perceptual, word-form area. *Journal of Experimental Psychology: General*, *131*(1), 65–72. https://doi.org/10.1037/0096-3445.131.1.65

Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *NeuroImage*, *193*, 167–177. https://doi.org/10.1016/j.neuroimage.2019.03.028

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865–7870. https://doi.org/10.1073/pnas.0509989103

Putnam, A. L., Ozubko, J. D., Macleod, C. M., & Roediger, H. L. (2014). The production effect in paired-associate learning: Benefits for item and associative information. *Memory & Cognition*, *42*(3), 409–420. https://doi.org/10.3758/s13421-013-0374-x

Qu, J., Pang, Y., Liu, X., Cao, Y., Huang, C., & Mei, L. (2022). Task modulates the orthographic and phonological representations in the bilateral ventral Occipitotemporal cortex. *Brain Imaging and Behavior*. https://doi.org/10.1007/s11682-022-00641-w

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*(8), 904–915. https://doi.org/10.1080/09658211.2013.766754

Quinlan, C. K., & Taylor, T. L. (2019). Mechanisms underlying the production effect for singing. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *73*(4), 254.

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), Article 1. https://doi.org/10.1038/nrn.2016.150

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.

Rice, G. E., Lambon Ralph, M. A., & Hoffman, P. (2015). The Roles of Left Versus Right Anterior Temporal Lobes in Conceptual Knowledge: An ALE Meta-analysis of 97 Functional Neuroimaging Studies. *Cerebral Cortex*, *25*(11), 4374–4391. https://doi.org/10.1093/cercor/bhv024

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife*, *5*, e18260. https://doi.org/10.7554/eLife.18260

Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. *Cerebral Cortex*, *23*(12), 2818–2828. https://doi.org/10.1093/cercor/bhs258

Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An Activation Likelihood Estimation meta-analysis. *Brain and Language*, *141*, 89–102. https://doi.org/10.1016/j.bandl.2014.11.012

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology*, *49*, 289–318. https://doi.org/10.1146/annurev.psych.49.1.289

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance*. *Bilingualism: Language and Cognition*, *15*(1), 157–166. https://doi.org/10.1017/S1366728910000623

Schmalz, X., Biurrun Manresa, J., & Zhang, L. (2021). What is a Bayes factor? *Psychological Methods*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/met0000421

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400–2406. https://doi.org/10.1093/brain/123.12.2400

Seghier, M. L. (2013). The Angular Gyrus. *The Neuroscientist*, *19*(1), 43–61. https://doi.org/10.1177/1073858412440596

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*(5), 238–242.

Seidenberg, M. S., Farry-Thorn, M., & Zevin, J. D. (2022). Models of word reading: What have we learned? In *The Science of Reading* (pp. 36–59). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119705116.ch2

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Smith, D. E., Moore, I. L., & Long, N. M. (2022). Temporal Context Modulates Encoding and Retrieval of Overlapping Events. *The Journal of Neuroscience*, *42*(14), 3000–3010. https://doi.org/10.1523/JNEUROSCI.1091-21.2022

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Squire, L. R. (2009). The legacy of patient HM for neuroscience. *Neuron*, *61*(1), 6–9.

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic Reinstatement in the Medial Temporal Lobe. *Journal of Neuroscience*, *32*(50), 18150–18156. https://doi.org/10.1523/JNEUROSCI.4156-12.2012

Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences*, *98*(22), 12760–12766. https://doi.org/10.1073/pnas.221462998

Taikh, A., & Bodner, G. E. (2016). Evaluating the basis of the between-group production effect in recognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 186.

Tanaka, S., & Kirino, E. (2021). The Precuneus Contributes to Embodied Scene Construction for Singing in an Opera. *Frontiers in Human Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnhum.2021.737742

Tange, O. (2011). Gnu parallel-the command-line power tool. *The USENIX Magazine*, *36*(1), 42–47.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*(11), 2119–2132. https://doi.org/10.1093/brain/122.11.2119

Taylor, J. S. H., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General*, *146*(6), 826.

Teichmann, L., Moerel, D., Baker, C., & Grootswagers, T. (2022). An empirically-driven guide on using Bayes Factors for M/EEG decoding. *Aperture Neuro*, *2*, 2021.06.23.449663. https://doi.org/10.1101/2021.06.23.449663

Thesen, T., McDonald, C. R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., Felsovalyi, O., Girard, H., Barr, W., Devinsky, O., Kuzniecky, R., & Halgren, E. (2012). Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nature Communications*, *3*(1), Article 1. https://doi.org/10.1038/ncomms2220

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words* (pp. xii, 274). Bureau of Publications, Teachers Co.

Tompary, A., Duncan, K., & Davachi, L. (2016). High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. *Hippocampus*, *26*(8), 995–1007. https://doi.org/10.1002/hipo.22582

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A Distributed Network for Multimodal Experiential Representation of Concepts. *Journal of Neuroscience*, *42*(37), 7121–7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

Trelle, A. N., Henson, R. N., & Simons, J. S. (2019). Neural evidence for age-related differences in representational quality and strategic retrieval processes. *Neurobiology of Aging*, *84*, 50–60. https://doi.org/10.1016/j.neurobiolaging.2019.07.012

Tremblay, P., & Gracco, V. L. (2009). Contribution of the pre-SMA to the production of words and non-speech oral motor gestures, as revealed by repetitive transcranial magnetic stimulation (rTMS). *Brain Research*, *1268*, 112–124. https://doi.org/10.1016/j.brainres.2009.02.076

Tremblay, P., & Gracco, V. L. (2010). On the selection of words and oral motor responses: Evidence of a response-independent fronto-parietal network. *Cortex*, *46*(1), 15–28. https://doi.org/10.1016/j.cortex.2009.03.003

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2011). Frontal pole cortex: Encoding ends at the end of the endbrain. *Trends in Cognitive Sciences*, *15*(4), 169–176. https://doi.org/10.1016/j.tics.2011.02.001

Tulkens, S., Sandra, D., & Daelemans, W. (2018). WordKit: A Python Package for Orthographic and Phonological Featurization. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 2695–2703.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology / Psychologie Canadienne*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Turkeltaub, P. E., Goldberg, E. M., Postman-Caucheteux, W. A., Palovcak, M., Quinn, C., Cantor, C., & Coslett, H. B. (2014). Alexia due to ischemic stroke of the visual word form area. *Neurocase*, *20*(2), 230–235. https://doi.org/10.1080/13554794.2013.770873

Umesh, P. (2012). Image processing in python. *CSI Communications*, *23*(2).

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine phonotactic online dictionary, Version 2.0.[Data file]*.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432. https://doi.org/10.1016/j.neuroimage.2005.11.002

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron*, *55*(1), 143–156. https://doi.org/10.1016/j.neuron.2007.05.031

Vogel, A. C., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2012). The Putative Visual Word Form Area Is Functionally Connected to the Dorsal Attention Network. *Cerebral Cortex (New York, NY)*, *22*(3), 537–549. https://doi.org/10.1093/cercor/bhr100

Wakeham-Lewis, R. M., Ozubko, J., & Fawcett, J. M. (2022). Characterizing production: The production effect is eliminated for unusual voices unless they are frequent at study. *Memory*, *30*(10), 1319–1333.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Wang, X., Xu, Y., Wang, Y., Zeng, Y., Zhang, J., Ling, Z., & Bi, Y. (2018). Representational similarity analysis reveals task-dependent semantic influence of the visual word form area. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-21062-0

Warrington, E. K., & Shallice, T. (1980). Word-form dyslexia. *Brain: A Journal of Neurology*, *103*(1), 99–112. https://doi.org/10.1093/brain/103.1.99

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, *97*(20), 11125–11129.

Wing, E. A., Ritchey, M., & Cabeza, R. (2015). Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. *Journal of Cognitive Neuroscience*, *27*(4), 679–691. https://doi.org/10.1162/jocn_a_00740

Woolnough, O., Forseth, K. J., Rollo, P. S., & Tandon, N. (2019). Uncovering the functional anatomy of the human insula during speech. *eLife*, *8*, e53086. https://doi.org/10.7554/eLife.53086

Xiao, X., Dong, Q., Gao, J., Men, W., Poldrack, R. A., & Xue, G. (2017). Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. *The Journal of Neuroscience*, *37*(11), 2986–2998. https://doi.org/10.1523/JNEUROSCI.2324-16.2017

Xue, G. (2018). The Neural Representations Underlying Human Episodic Memory. *Trends in Cognitive Sciences*, *22*(6), 544–561. https://doi.org/10.1016/j.tics.2018.03.004

Xue, G. (2022). From remembering to reconstruction: The transformative neural representation of episodic memory. *Progress in Neurobiology*, *219*, 102351. https://doi.org/10.1016/j.pneurobio.2022.102351

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*(6000), 97–101.

Xue, G., Dong, Q., Chen, C., Lu, Z.-L., Mumford, J. A., & Poldrack, R. A. (2013). Complementary Role of Frontoparietal Activity and Cortical Pattern Similarity in Successful Episodic Memory Encoding. *Cerebral Cortex*, *23*(7), 1562–1571. https://doi.org/10.1093/cercor/bhs143

Ye, Q., Zou, F., Lau, H., Hu, Y., & Kwok, S. C. (2018). Causal evidence for mnemonic metacognition in human precuneus. *Journal of Neuroscience*, *38*(28), 6379–6387.

Zeithamova, D., de Araujo Sanchez, M.-A., & Adke, A. (2017). Trial timing and pattern-information analyses of fMRI data. *NeuroImage*, *153*, 221–231. https://doi.org/10.1016/j.neuroimage.2017.04.025

Zeller, D., Litvak, V., Friston, K. J., & Classen, J. (2015). Sensory Processing and the Rubber Hand Illusion—An Evoked Potentials Study. *Journal of Cognitive Neuroscience*, *27*(3), 573–582. https://doi.org/10.1162/jocn_a_00705

Zhang, B., Meng, Z., Li, Q., Chen, A., & Bodner, G. E. (2023). EEG-based univariate and multivariate analyses reveal that multiple processes contribute to the production effect in recognition. *Cortex*, *165*, 57–69. https://doi.org/10.1016/j.cortex.2023.04.006

Zhang, W., Liu, Y., Wang, X., & Tian, X. (2020). The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. *Cognitive Neuroscience*, *11*(4), 194–204. https://doi.org/10.1080/17588928.2020.1792868

Zhao, L., Chen, C., Shao, L., Wang, Y., Xiao, X., Chen, C., Yang, J., Zevin, J., & Xue, G. (2017). Orthographic and Phonological Representations in the Fusiform Cortex. *Cerebral Cortex*, *27*(11), 5197–5210. https://doi.org/10.1093/cercor/bhw300

Zhou, Y., & MacLeod, C. M. (2021). Production between and within: Distinctiveness and the relative magnitude of the production effect. *Memory*. https://www.tandfonline.com/doi/full/10.1080/09658211.2020.1868526