Summer 8-15-2019

Isolating Item and Subject Contributions to the Subsequent Memory Effect

Jihyun Cha
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Psychological and Brain Sciences

Dissertation Examination Committee:
Ian G. Dobbins, Chair
David A. Balota
Kathleen B. McDermott
Jonathan E. Peelle
Henry L. Roediger, III

Isolating Item and Subject Contributions to the Subsequent Memory Effect

by

Jihyun Cha

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor, Ian G. Dobbins for his untiring support and patience throughout my academic development. I'd also like to thank the members of the Dobbins lab, past and present, for being such great colleagues and friends. I thank my committee for their invaluable insight and feedback throughout the entire dissertation process. I also want to thank the scholars and the staffs of the McDonnell International Scholars Academy. It was a great honor to be a part of such an inspirational group. Also, I would like to thank the ambassador to my Alma Mater, Drs. Jin-Moo Lee and Katie Vo, for being amazing hosts and mentors. Finally, I thank my dearest friends and family for always being there for me.

그리고 어머니, 아버지, 끊임없는 사랑과 노고에 감사드립니다.

This research was supported by an American Psychological Association Dissertation Research Award.

Jihyun Cha

Washington University in St. Louis

August 2019

ABSTRACT OF THE DISSERTATION

Isolating Item and Subject Contributions to the Subsequent Memory Effect

by

Jihyun Cha

Doctor of Philosophy in Psychological and Brain Sciences

Washington University in St. Louis, 2019

Ian G. Dobbins, Chair

The subsequent memory effect (SME) refers to the greater brain activation during encoding of subsequently recognized items compared to subsequently forgotten items. Previous literature regarding SME has been primarily focused on identifying the role of specific regions during encoding or factors that potentially modulate the phenomenon. The current dissertation examines the degree to which this phenomenon can be explained by item selection effects; that is, the tendency of some items to be inherently more memorable than others. To estimate the potential contribution of items to SME, I provided participants a fixed set of items during encoding, which allowed me to model item-specific contributions to recognition memory strength ratings using a linear mixed effect (LME) model. Using these item-based estimates, I was then able to isolate two distinct item-related activations during encoding that were linked to item distinctiveness and general item memorability, respectively. However, the residual of the LME model which reflects recognition strength unaccounted for by the items recovered the majority of original areas linked

to subsequent recognition. Thus, I conclude that SMEs are largely attributable to encoding-related processes unique to each subject. Nevertheless, proper modeling and statistical control of item-driven effects afforded detection of originally missed encoding-activations and resulted in a SME more robust than the original. Taken together, these findings suggest that the SME reported in the literature is largely independent of the specific items encoded and demonstrates the need for different functional interpretations of item- versus subject-driven SMEs.

Chapter 1: Introduction

One of the most studied memory phenomena in functional brain imaging is the subsequent memory effect (SME). Beginning with work by Wagner et al. (1998) and Brewer et al. (1998), it was noted that during encoding, words that later go on to be recognized yield higher activation than those later forgotten in several brain regions (Figure 1.1) including the left inferior frontal cortex (IFC), bilateral fusiform cortex, bilateral hippocampal formation, bilateral premotor cortex (PMC) and bilateral posterior parietal cortex (PPC) (Kim, 2011). These findings evoked considerable interest because they suggested the possibility of capturing the online process of encoding or at least those processes leading to durable encoding.

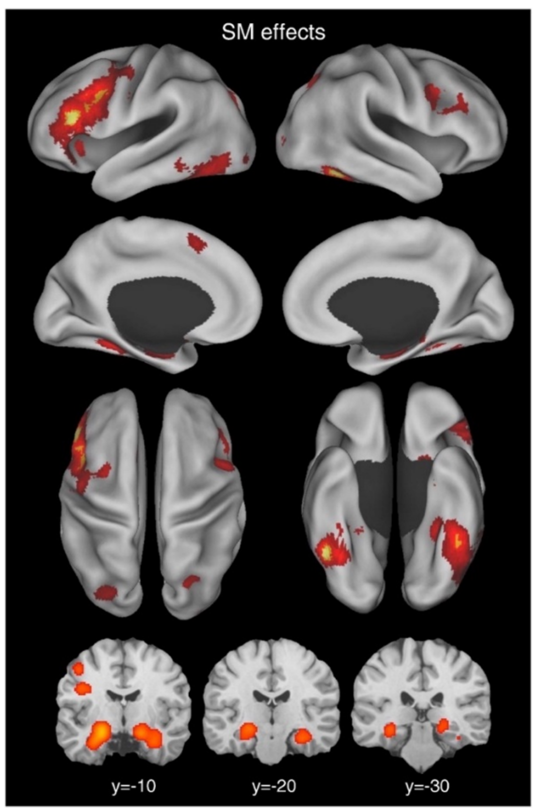

Figure 1.1. Subsequent memory map (adapted from Kim, 2011)

As in most cognitive neuroimaging research, studies on the SME have been primarily focused on identifying the functional roles of specific regions during encoding. For example, SMEs in the medial temporal regions support the general idea that ‘strong’ activation of these memory-linked areas promotes durable encoding (Brown & Aggleton, 2001; Davachi, 2006; Davachi, Mitchell, & Wagner, 2003; Eichenbaum, Yonelinas, & Ranganath, 2007; LaRocque et al., 2013; Liang, Wagner, & Preston, 2012; Squire & Zola-Morgan, 1991). Another line of research has been focused on the factors that potentially modulate the phenomenon. For example, how the composition of regions demonstrating SME varies via stimulus class or type of memory assessed (Kim, 2011; Kirchhoff, Wagner, Maril, & Stern, 2000; Qin, van Marle, Hermans, & Fernández, 2011; Uncapher & Wagner, 2009), whether there is an overlap between the regions demonstrating SMEs and the regions responding to memory relevant item characteristics such as word frequency (Chee, Westphal, Goh, Graham, & Song, 2003) and finally, whether the magnitude of SMEs in certain regions of interest (ROIs) changes as a function of factors such as phonological familiarity (Clark & Wagner, 2003), arousal (Dolcos, LaBar, & Cabeza, 2004), or aging (Daselaar, Veltman, Rombouts, Raaijmakers, & Jonker, 2003).

Common to most of these approaches, is an interest in identifying cognitive processes that affect encoding that **generalize** across exemplars of the particular stimulus categories tested. In an attempt to achieve this generalizability, the materials are counterbalanced across conditions of interest. For example, in a study of cognitive aging and verbal recognition memory, one would counterbalance the words across old and new study status, with both young and older adults receiving the same counterbalancing. However, because the subjects’ performance at the level of items determines the two conditions of primary interest (recognized versus forgotten) the counterbalancing methods, in isolation, can never ensure that the observed activation differences are not largely the result of simply comparing more versus less intrinsically memorable items. In the case of cognitive aging, this would be potentially confounded by the possibility that this potential item selection artifact could differ across cohorts in ways unrelated to encoding processes per se. For example, if the vocabularies markedly differ across the two groups in a way systematically linked to the characteristics of the items, then SME differences could arise even if the effect within each group largely reflected an item selection effect.

Thus, while a given researcher might conclude that a particular region of activation difference in the above example reflect, for example, differences in self-initiated semantic elaboration strategies that lead to differences in subsequent memory, they could instead reflect the differences between intrinsically more, versus less memorable items. Indeed, as I further discuss below, item selection effects in SME designs are pernicious because they could arise even if each subject receives a wholly different set of items. All that is required is that one assume that items differ intrinsically in memorability. The key challenge posed by these potential item selection effects, is that they compete with other process-oriented explanations that should span items per se (e.g., self-initiated elaborative strategies). Below I discuss the origins of the SME design in more detail, highlighting the potential for item selection effects within the designs.

1.1 Challenging Aspects of the Subsequent Memory Paradigm: Self-selection by subject

While motivated by similar phenomena discovered from ERP studies (Paller, Kutas, & Mayes, 1987), examination of SMEs with functional magnetic resonance imaging (fMRI) was limited because of the slow temporal dynamics of the BOLD signal, which made it difficult to link the response to an event during a specific trial. Thus, the SME in fMRI research was originally studied by comparing the two blocks where the participants engage in the different encoding processes known to produce higher versus lower level of recollection (e.g., semantic versus non-semantic encoding) (See Wagner et al., 1998 for the discussion of this issue). Only after the development of event-related designs (Dale & Buckner, 1997), was the trial-wise examination of encoding activations made possible via fMRI. Rapidly after this methodological advance, two seminal studies ported the SME design of ERP research into fMRI event-related designs to document spatially precise SMEs using verbal items (Wagner et al., 1998) and pictorial items (Brewer et al., 1998).

Even with this methodological advance, which allowed researchers to spatially localize specific regions demonstrating the SME pattern for specific encoding conditions, the interpretation of the SME activations critically depends upon whether one assumes items are equally capable of being encoded. If the assumption is true, then the contrast of subsequent hits versus misses reveals the recruitment of subject-initiated processes, such as semantic elaboration or attentional focus, which facilitate memory encoding. If not however, then the contrast of hits versus misses may merely reveal that some items are intrinsically more memorable than others. This sort of item selection effect, if present, would challenge interpretations that depend upon subject-initiated encoding processes, and instead suggest more passive encoding interpretations linked to item properties such as distinctiveness.

To some extent, the failure to more seriously consider item selection interpretations in prior SME research may have reflected the fact that items are invariably counterbalanced across old and new stimulus categories, perhaps leading to the tacit belief that item selection effects are therefore minimized. However, the efficacy of counterbalancing is completely negated by the fact that subject-selections define the categories of interest and the researchers have no control over which item to be later remembered or forgotten. To see why, consider the possibility that all items have a normative probability of being well-encoded (viz., item memorability) that ranges from 0 to 1. If this were true, the same selection effect would occur across participants even if each participant **studied an entirely different set of items**. Then, his or her performance would merely reflect (or be conflated with) the a priori memorability of the items or rank ordering across items that he or she was given to study. Thus, as I show below, the only way to effectively deal with potential item selection effects, is to directly model their contribution to performance. Perhaps, somewhat surprisingly, the simplest way to do so is to give every subject the same set of items during encoding.

To my knowledge, there has not been a single functional imaging study of verbal recognition SMEs that has attempted to model the contribution of item selection effects to the SME findings (although see Bainbridge, Dilks, & Oliva (2017) for a pictorial recognition work). However, behavioral researchers have been interested in the contributions of various normative word characteristics that might lead to increased versus decreased recognition. For example, the work of Cortese and colleagues (2010; 2015) which I discuss more fully in Section 1.2 might be construed as demonstrating that item selection effects in recognition might be quite large and those effects are linked to easily identifiable psycholinguistic characteristics of the tested words. Briefly, they demonstrated that items can differ in their consensual memorability (probability of recognition across individuals) and that a substantial amount of variance for these memorability scores can be explained as a function of normative item characteristics such as word frequency and imageability. These studies raise the possibility that some part of the SME documented in prior functional imaging research may simply reflect the operation of these item-level characteristics rather than subject-initiated processes that govern encoding efficacy (e.g., self-initiated semantic elaboration).

Another demonstration of potentially large item selection effects comes from a recent methodological paper from Westfall, Nichols and Yarkoni (2016). The researchers' goal was to illustrate the large role that stimulus effects can play in statistical inference if incorrectly modelled as a fixed effect (Clark, 1973). Through simulation and re-analysis of extant data, they demonstrated that modeling stimulus-level variability in fMRI designs tends to markedly reduce the effect sizes attributed to contrasts of conditions across those stimuli. Specifically, to demonstrate the magnitude of the 'stimulus-as-fixed-effect fallacy' in fMRI, they resorted to one of the most well-established findings which is the role of the amygdala in affective processing as an example. With the standard model assuming the equal subject's response to the stimuli within each condition, the "emotional faces" (10 exemplars in each emotional class; anger and fear) produced a robust amygdala response compared to "geometric shape" (3 exemplars) and a smaller but notable increase for "anger faces" relative to "fear faces". However, even with 111 participants in the analyses (which is an unusually large sample for typical fMRI studies), the revised model assuming random stimulus variability demonstrated an 89% reduction in the effect of face vs. shape contrast and 78% reduction in the anger vs fear contrast rendering latter ambiguous. Interestingly, Westfall, Nichols and Yarkoni (2016) also demonstrated that the concern over item variability may be reduced for designs with large numbers of items in the respective classes. This presumably reflects the fact that across-item variability in evoked activation becomes increasingly less critical with large numbers of items in each class (i.e., the class means are estimated more reliably). Nonetheless, it is important to note that the examples considered in the study did not have a subject-selection component, and hence the findings do not directly inform concerns about the interpretation of SMEs in fMRI where each subject determines which items fall into the classes of interest. In relevance to the current dissertation, Westfall and colleagues also introduced a secondary benefit of modeling random stimulus effect, asserting that the inclusion of a separate parameter for each stimulus allows one to estimate the brain activation associated with each stimulus. This is consistent to the methodological framework of the current dissertation introduced in section 3.2.4, where I estimate the general memorability of each item using linear mixed effect (LME) modeling approach and then incorporate these estimates as a new parametric modulator to detect a new set of regions associated with item effects during encoding.

The item-focused analyses that I consider in this dissertation can be divided into two conceptually distinct approaches based on whether item encoding influences are modeled as resulting from normative psycholinguistic characteristics (normative characteristics approach), or instead estimated via the aggregate performance of other subjects (item memorability approach).

1. Normative characteristics approach. Analogous to studies in psycholinguistics, I explore several normative word characteristics, looking for regions sensitive to trial-wise variation in these characteristics during encoding. I then compare these activation maps to the SME map defined in the traditional manner by contrasting subsequent hits and misses. The degree to which the trial-wise effects of the normative word characteristics overlap the traditional SME, is important because it suggests, albeit informally, how concerned one should be that the traditional SME may reflect a property of normative item characteristics as opposed to subject-initiated cognitive processes that transcend the specific items being encoded (i.e., subject-driven effects).

As noted, this type of research has been primarily conducted in psycholinguistics (Kronbichler et al., 2004; Pexman, Hargreaves, Edwards, Henry, & Goodyear, 2007; Schuster, Hawelka, Hutzler, Kronbichler, & Richlan, 2016; Yarkoni, Speer, Balota, McAvoy, & Zacks, 2008). These studies have used parametric modulation to model the trial-wise fluctuation in brain response in accordance with the fluctuation in each variable. For example, Yarkoni et al. (2008) and Schuster et al. (2016) reported that activation in the putative visual word forma area (VWFA) showed reliable (negative) correlations with lexical variables such as word frequency, suggesting that, in contrast to the original conceptualization of the region as an area dedicated to pre-lexical, perceptual processing of word forms (Cohen et al., 2002; Dehaene, Cohen, Sigman, & Vinckier, 2005; Dehaene et al., 2002; McCandliss, Cohen, & Dehaene, 2003), its role extends to more abstract representations of words (also see Price & Devlin, 2003; Vogel, Petersen, & Schlaggar, 2014 for the region's involvement beyond the orthographic features of words). Similarly, one goal of this dissertation is to detect regions demonstrating the item-level modulation from the normative item variables and to see if they overlap with regions demonstrating SME.

2. Item memorability approach. The second approach uses item memorability estimates, typically calculated as a simple consensual hit rate of the group as the focus of the analysis. Consensuality measures simply record the tendency of observers to respond in the same manner for each item (Koriat, 2008), for example, item memorability measured by consensual hit rate of Cortese et al. (2015) described below. Given that the scope of this dissertation is limited to recognition memory and the SMEs measured by recognition test, by the term “memorability”, I refer to recognition memorability and its relationship to normative variables throughout the dissertation.

In the case of Cortese et al. (2015) the goal was to use normative item characteristic in an attempt to explain variation in item memorability scores. A new question I will address in the current dissertation instead examines the degree to which consensual memorability scores predict individual subject behavior across trials. This question directly examines the degree to which each subjects' responding is, or is not, predictable as a function of the response tendencies of others to the same items. As I will discuss in following sections, it is possible to explain a considerable amount of variation in item memorability scores (using, for example, normative characteristics), even if those memorability scores are not robust predictors of recognition decisions across the trials of observers.

To summarize, I will focus on the two types of item measures, in order to examine: (1) How well do the normative characteristics of items (e.g., word frequency) explain variation in either consensual memorability scores, or the efficacy of encoding across items? (2) Regardless of these characteristics, to what degree can we predict the memory responses of an individual, by knowing the consensus responses of others to the items that he or she is being tested on?

To provide more wholistic picture of item contributions in SME, I will combine these item measures within a single LME equation predicting trial-wise subsequent recognition memory strength reports. The predictions arising from these two separable constituents of the LME model will then be used to parametrically model brain activation during encoding. This LME model prediction approach directly addresses the primary question of the dissertation, the degree to which traditional SMEs during verbal encoding are subject- versus item-driven phenomena. If they were fully the former, then item-based prediction of recognition strength reports would not implicate regions identified through the traditional SME contrast (hits > misses). If the traditional SMEs were fully an item-driven phenomenon, then the item-based LME model prediction would largely, if not fully, encompass the traditional SME map. In other words, the SME would largely reflect item variations in intrinsic memorability.

Below I take a closer look at behavioral studies suggesting a large item-level correspondence between normative lexical variables and recognition memory.

1.2 Normative Item Characteristics Linked to Recognition Memory

Extending a previous 'mega-study' approach of lexical decision-making to recognition memory, Cortese and colleagues examined the relationship between several potential item-level characteristics (Cortese, Khanna, & Hacker, 2010; Cortese, McCarty, & Schock, 2015; Lau, Goh, & Yap, 2018) and the consensual hit rate measured for each item (i.e., item memorability).

In their recent study using 2,897 disyllabic words (Cortese et al., 2015), the authors demonstrated the relative contribution of each normative variable using the multiple regression framework. Remarkably, these variables jointly accounted for about 35% of the total variance in consensual hit rates as well as the difference between the consensual hit and false alarm rates for the items. Table 1.1 summarizes the list of the item variables used in the study and the standardized regression coefficients of the variables from the multiple regression model predicting consensual hit rates.

Table 1.1. Standardized regression coefficients in multiple regression model predicting consensual Hit rates at the item-level. (Adapted from Cortese et al., 2015)

Factor	*Hits*
Imageability	.487**
AoA	.323**
Length	−.163**
Log frequency	−.258**
LOD	.099**
LPD	−.027
R^2	.350**

Note: ** $p < .01$.

*AoA = age of acquisition; LOD = Levenshtein orthographic distance; LPD = Levenshitein phonological distance.

As discussed earlier, in explicit memory research (not limited to studies on SME) the assessment of item-level effects is infrequent. In accordance with the first, item characteristics approach, I will examine whether the findings of Cortese et al. (2015) above generalize to a different encoding task and whether the normative word characteristics identified by the authors explain SMEs in fMRI data. Additionally, I will compare the memorability scores of items across the Cortese study and my current sample, to see if consensuality-based item memorability scores are stable and predictive of SMEs. Critically, if the memorability scores derived from one set of subjects explain the item memorability effects in a separate group of subjects, it necessarily implies that different individuals possess similar mental representations of items that are germane to encoding and retrieval of outcomes. Otherwise such regularities could not be observed. Bainbridge and colleagues (Bainbridge et al., 2017; Bainbridge & Rissman, 2018) pushed this idea to the extreme and argued that memorability itself is intrinsic property of each item. I will introduce the findings from these studies below and discuss some of their claims that I tested in the following sections.

1.3 Memorability as Intrinsic Item Property

As mentioned above, Bainbridge and colleagues studied "item memorability" as their primary variable of interest arguing that memorability tags the statistical distinctiveness of stimuli for later encoding. Based on their interest in the question of how perceptual processing of visual stimuli (e.g., images for faces and scenes) progresses to encoding, they collected memorability scores of 720 images from 800 observers outside the study. They claimed that memorability can be used as an intrinsic attribute of the images based on the demonstration that the ranking of the images in terms of their memorability is consistent across participants (The random split-half reliability –average correlation between the memorability scores based on the first half of participants and that of the second half- was .69 for faces and .75 for scenes indicated by Spearman's rank correlation ρ).

Although their memorability scores (consensual hit rate for each item) were in a continuous scale, in their fMRI study on encoding (Bainbridge et al., 2017), they dichotomized this variable (memorable versus forgettable images) presumably to match the subsequent memory contrast also conducted on these data (his > misses). Both of their univariate and multivariate analyses comparing the two constructs (memorability versus subsequent memory) consistently demonstrated that neural substrates of memorability were dissociable from those of individual participant's subsequent memory. Specifically, memorability effects were found in the ventral visual stream (VVS) and the medial temporal lobe (MTL) whereas SMEs were in the prefrontal cortex (PFC). Based on the dissociation between two apparent "memory" related constructs, they argued that memorability is a stimulus property that is intrinsic to an item which bridges the gap between perception and memory.

The claim that memorability is "intrinsic" property of the stimuli per se might be reasonable for the perceptual stimuli that Bainbridge and colleagues used. However, it raises several questions. First, the degree to which the memorability contrast should overlap with the SME contrast, presumably depends on how much memorability (consensually defined) explains the recognition decisions of each individual. If memorability scores only explain a small (but systematic) proportion of each subject's subsequent recognition decisions, then it seems unlikely that, the two maps would be coincident. Second, whether the memorability of verbal items demonstrates such consistency across observers (the very finding that led the authors to conclude memorability for images is an intrinsic stimulus property) is unclear. For example, verbal items may be less dependent on the "perceptual" or structural qualities of each word (although orthographical & phonological aspects may be important) and more dependent on the situation/contexts of encoding and retrieval. In fact, numerous studies have demonstrated the importance of encoding operations (e.g., levels-of-processing manipulation) or the match between encoding and retrieval contexts (e.g., encoding specificity principle and transfer-appropriate processing) on verbal memory (Craik & Lockhart, 1972; Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977; Thomson & Tulving, 1970; Tulving & Thomson, 1973), both of which suggest that the items that are memorable within one context or task might not be equally memorable within another. If relative ranking of memorability across items is heavily altered via these manipulations, then the claim that memorability is an intrinsic item property is tenuous. This said, there do not appear to be studies that specifically look at the rank orderings of item memorability estimates under different manipulations or contexts with a goal towards showing that memorability is more or less stable across contexts for certain types of materials. For example, as mentioned earlier, it is currently unclear whether the ordering of the Cortese et al. (2015) memorability estimates would substantially change across specific encoding tasks or manipulations of encoding and retrieval contexts. In this regard, it is important to distinguish between the fact that a manipulation can substantially lower or increase net recognition and that it can substantially alter the rank ordering of items in terms of consensual memorability.

To sum up, whether the memorability of verbal items should be viewed as an intrinsic property (independent of, say, encoding operations) and whether the brain regions sensitive to verbal item memorability estimates are distinct from traditionally defined SME regions, are empirical questions to be tested. To address these, memorability scores taken from Cortese as well as calculated from the current study will be examined.

1.4 Distinction between Item- and Trial-level Analysis: Explaining item effects across versus within observers.

As discussed earlier, the item characteristics approach tries to explain variation in consensual memorability scores using normative item variables such as word frequency or imageability. This is conceptually different from explaining the recognition performance of an individual across trials by consensual memorability and/or by normative item characteristics which may also operate at the level of each trial. Hence, in actual analyses, the same explanatory variable such as normative item characteristics can be used to explain both trial-level performance and item-level memorability estimates. Thus, it is important to highlight whether one is modeling variation across aggregated scores for items (item-level) versus trying to explain variation across the trials within observers (trial-level). For example, in Cortese et al. (2015), every measurement is collapsed or summarized across subjects for each item so that performance of each individual is no longer observable.

Even if one finds a set of normative characteristics that explains a considerable proportion of variation in consensus scores for an item-level analysis (as in Cortese et al.), this does not mean that the consensus scores themselves explain a particularly large amount of variation at the level of trial outcomes within individuals. This is because consensus measures are aggregates and they can lead to stable relationships with normative variables even if that variable only has a minor effect on the trial-wise behavioral outcomes of each individual. In this regard, it is important not to conflate the size of normative item characteristic's contribution in explaining consensuality data (item-level) with the ability of that variable to explain variation across the recognition trials within individuals (trial-level).

To claim that an individual's recognition memory behavior is heavily determined by item effects broadly, is to claim that his or her performance can be well predicted by knowing the specific items that were studied. This in turn means that his or her performance would be highly predictable by knowing the consensus response of others to those same items, if it were the normative features of items that produce the effects. Again, knowing how much variance in the consensus (item memorability) can be accounted for by normative item variables (as in Cortese et al.) does not inform us in predicting each participant's memory performance across trials, except perhaps at the boundary conditions of perfect consensus (i.e., every item has a group hit probability of 1 or 0) or no consensus whatsoever (i.e., every item has a group hit probability .5). In the former case, we should expect to explain all the variation in each subject's recognition performance, whereas in the latter case, we should expect to explain none of the variance. Outside of these extremes, the question of the links between explaining variation across consensus measures, versus using consensus measures to explain variation in the trials within individuals, is empirical. In this dissertation I will demonstrate that the normative item characteristics identified by Cortese et al., and the consensus item memorability estimates of that report (or calculated from the current participants) do not explain much of the recognition behavior at the level of trials within participants.

1.5 Research Aims

In order to examine potential item contribution within SMEs, I will explicitly model the items by providing a fixed list of items to the participants. This will allow me to address both levels (item versus trial) of item contribution.

First, with item-level analyses, I will examine the effects of normative characteristics on consensus in accuracy for each item (i.e., item memorability as dependent variable). Findings from this analysis explore why, in the aggregate, certain items are more memorable than others. If one or some combination of normative item characteristics can explain a decent amount of variance in memorability of verbal items (as shown in Cortese et al. 2015) and if memorability is implicated in the regions demonstrating SMEs (despite the dissociation between memorability map and SME map for "pictorial items" reported by Bainbridge et al., 2017), we will be able to observe a decent overlap between regions sensitive to those normative contributors and regions showing SMEs.

Also, as an important first step in considering the stability of consensus measures of memorability, which must necessarily be high if memorability is an intrinsic property of each item, I will compare the consensus hit rates from my sample, which will use deep processing for encoding, to those of Cortese, who used unstructured encoding. Critically, if consensus hit rates were extremely similar across the Cortese and current data, and if these hit rates predicted encoding activations that strongly overlap with those defined using the traditional SME contrast, it would indicate that SME effects are strongly a function of normative item memorability. Also, because Cortese et al. were able to account for more than 35% of the variation in consensus hit rates using specific normative word characteristics, we would gain insight into the mechanisms driving the memorability of verbal items.

Second, based on item effects analysis, I will address the degree to which the behavior of individuals at the level of trials can in fact be explained using consensus memorability estimates. Although the literature warning against "item-as-fixed-effect fallacy" (Clark, 1973; Westfall et al., 2016) can be taken to suggest the item effects in recognition memory are quite large, as I have noted above, this may reflect confusion with the fact that the reliability of consensuality measures is large. Thus, I will directly test how effective consensuality measures are in predicting trial-level outcomes across individuals. Behaviorally, if there are considerable item effects in recognition memory, we will be able to make prediction on one's recognition performance in each trial just by knowing how others have responded to the item. In imaging, regions and their magnitude of activation detected by individual's memory performance (SME) should be comparable to those detected by group tendencies. Following recommendations of studies emphasizing random stimulus effects (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013), an LME modelling framework will be adopted to capture the random effect of items and examine the item contributions in the trial-wise recognition behavior of individuals scanned during encoding. Components of the model will then be used to isolate SMEs that either are, or are not, reliably linked to item effects. Finally, the overall utility of the estimates from the framework in imaging analysis will be also discussed.

To preview, the data suggest that 1) consensuality measures of verbal memorability and its relationship to item characteristics may not be highly stable across differences in context, 2) item effects in behavioral recognition performance are quite small, and 3) the SME in functional imaging largely reflects subject-driven processes that cannot be explained by knowing which particular items a subject is encoding.

Chapter 2: Experiment

2.1 Participants

Twenty-five participants from Washington University in St. Louis and the St. Louis community were recruited. Participants were right-handed, native English speakers with normal or corrected-to normal vision and no history of psychiatric or neurological illness (via self-report). The participants received $25 per hour (up to three hours) as compensation. All participants provided informed consent in accordance with university guidelines.

Two participants were excluded due to poor recognition performance (*d*-primes of .05 and .37 respectively, with the latter also failing to respond on 188 of 400 encoding trials). Another participant was excluded due to severe image artifact (no signal at the superior parts of the brain), leaving 22 for the analyses. The remaining participants were 18 – 29 years old (M = 21.5, SD = 2.96) and nine were female.

2.2 Materials

All the tasks were performed on an IBM laptop running PsychoPy (version 1.85.4; Peirce, 2009). During scanning, stimuli were presented via an MR-compatible rear projector (screen resolution of 1024 x 768 pixels) viewed through a mirror attached to the head coil. All encoding items were presented centrally in white on a black background. In-scanner responses were made via button press on a response box.

Outside the scanner, test stimuli were displayed on the built-in display of the laptop (screen resolution of 1366 x 768 pixels) and the responses were collected via keyboard.

2.2.1 Word list preparation

600 nouns (400 recognition targets and 200 lures) were drawn from the 2,897 disyllabic words of Cortese et al. (2015). The 1,237 available nouns were first sorted into 10 levels in terms of their memorability (consensual hit rate) with 60 items randomly drawn from each decile, ensuring the resulting 600 items would span a representative range of normative memorability. This random selection process was repeated 2,000 times, and the one yielding the strongest association between semantic distinctiveness (defined below) and normative memorability was retained as the final list for the experiment, maximizing the potential for observing the mediating effects of semantic distinctiveness during encoding. Finally, 400 out of 600 selected nouns were randomly assigned as the fixed list of encoding materials, with the remaining 200 serving as lures during the post-scanning recognition test.

The original Cortese list was normed for various attributes such as Imageability (originally taken from Schock, Cortese, Khanna, & Toppi, 2012; a 1-7 scale, rated by 30 subjects; **IMG** for short), **Length**, Word frequency (Brysbaert & New, 2009; subtitle norms, the common log of the frequency per million words estimate; here, I inversed this value [1/word frequency] so that its relationship to memorability matches to other variables; **InvLogFreq**), Age of Acquisition (taken from Schock, Cortese, Khanna, and Toppi, 2012; subjective estimate of age range in a 1-7 scale, 1 indicating age between 0 and 2, 7 indicating age of 13 or later, rated by 32 subjects; **AoA**), and finally, Orthographic Distinctiveness (measured by Levenshtein orthographic distance; **Ortho Dist**), and Phonological Distinctiveness (measured by Levenshtein phonological distance; **Phono Dist)**[1].

An additional characteristic considered here was Semantic Distinctiveness (**Semantic Dist**) of the items (The details of the vector smantics approach used to produce the variable are described in Appendix I). To quantify semantic distinctiveness of the words, I examined the pattern of occurence of each word across a large corpus (*N* = 100, 000) of movie reviews (Maas et al., 2011). Critically, words that have a unique distiribution of occurrence across the reviews (relative to the remaining words) were assumed to be semantically distinctive. While treating the 100,000 movie reivews as an n-dimensional space in which words could cluster or isolate in terms of their relative positions (determined by the similarity of the occurrence pattern) I first calcuated pairewise cosine similarity of each word with respect to every other word in the set (the entire 2,897 disyllabic words from Cortese et al., 2015). The similarity value for each word was then translated into a dissimilarity (viz., a distinctiveness score) by subtracting it from 1, the highest possible score of similarity. Finally, the mean of these scores was calculated for each word so that the mean value can be used as a normative variable linked to the word. A high mean indicates that the word was, on average, distributed distinctively across the movie reviews presumably due to its distinctive meaning in comparison with the remainder of the set. Note that the set of words used for the computation of the pairwise cosine similarities included the entire Cortese disyllabic word list. Thus, the resulting semantic distinctiveness measure is the word's

---

[1] The Levenshtein distances are computed using the number of deletions, substitutions and insertions necessary to change one letter string into another, with distinctiveness calculated as the mean Levenstein distance for each word to its closest 20 neighbors. Words with higher value are less similar to others (= more distinctive). (Yarkoni, Balota, & Yap, 2008).

relative distinctiveness within a database of 2,897 words with a wide range of grammatical classes, not limited to the selected 400 nouns. This helps ensure the measure is a “normative” variable rather than a study-specific variable. Table 2.1 summarizes the seven normative item charactersitics considered in the study and Table 2.2 shows the correlations among these characteristics for the 400 encoding items.

Table 2.1. Descriptive statistics of seven normative variables for the encoding words ($N = 400$)

	IMG	AoA	Length	InvLogFreq*	Ortho Dist	Phono Dist	Semantic Dist
Mean	5.084	4.955	6.155	.469	2.300	2.159	.998
SD	1.298	.915	1.122	.110	.498	.496	.003

* Note that the Log Word Frequency was inversed (1/WF) to match the direction of other measures, such that a larger value (potentially) leads to more distinctive processing and better encoding.

Table 2.2. Correlation matrix of seven normative variables used in the study.

	IMG	AoA	Length	InvLogFreq	Ortho Dist	Phono Dist	Semantic Dist
IMG	-	-0.61***	-0.05	-0.34***	-0.05	-0.14**	0.08
AoA		-	0.07	0.58***	0.18***	0.19***	0.32***
Length			-	0.02	0.67***	0.49***	-0.03
InvLogFreq				-	0.15**	0.16**	0.61***
Ortho Dist					-	0.61***	0.08
Phono Dist						-	0.05

Computed correlation used pearson-method with listwise-deletion.

2.2.2 Nonword list preparation

One hundred pronounceable nonwords whose lexical characteristics (number of syllables, length, bigram frequency characteristics and the lexical decision performance measures) were matched to those of the encoding list were generated from the English Lexicon Project (ELP: Balota et al., 2007) website (http://elexicon.wustl.edu/NonWordStart.asp). Nonwords were included in the current design as a potential interpretive check on subsequent memory findings. For example, if a region demonstrated an SME was also linked to semantic distinctiveness, the interpretation that semantic distinctiveness mediated the SME finding would also require that the region distinguishes between words and nonwords, since the latter would have minimal semantic contributions.

2.3 Procedure

The experiment consisted of three major phases: incidental encoding within the scanner, recognition testing outside the scanner and finally, two short verbal intelligence tests also conducted outside the scanner.

After being situated in the scanner, the participants first underwent the structural scanning. This lasted nine minutes for five participants and five minutes for the remainder due to switching the scanning protocol after the first five participants to shorten scanning time using a parallel acquisition technique for the structural images. Functional images were collected in four separate scans constituting four blocks of the encoding task.

During encoding, participants reported whether the presented word was “pleasant” or “not pleasant”. If given a nonword, they were told to report the item as “not pleasant” even if the presented nonword resembles or reminds them of some pleasant real word. If a fixation cross was presented on the screen, the participants waited for the next item to appear while remaining focused. Stimuli were presented for four seconds during which responses were made via a MR-compatible response box. Two buttons on the response box were used, each corresponding to YES and NO decisions respectively to the question “Is this one pleasant?”. YES was indicated using the index finger and NO the middle finger. A brief break was given after finishing each block of 125 judgment trials (100 words and 25 nonwords). The entire in-scanner encoding phase took approximately an hour for each participant.

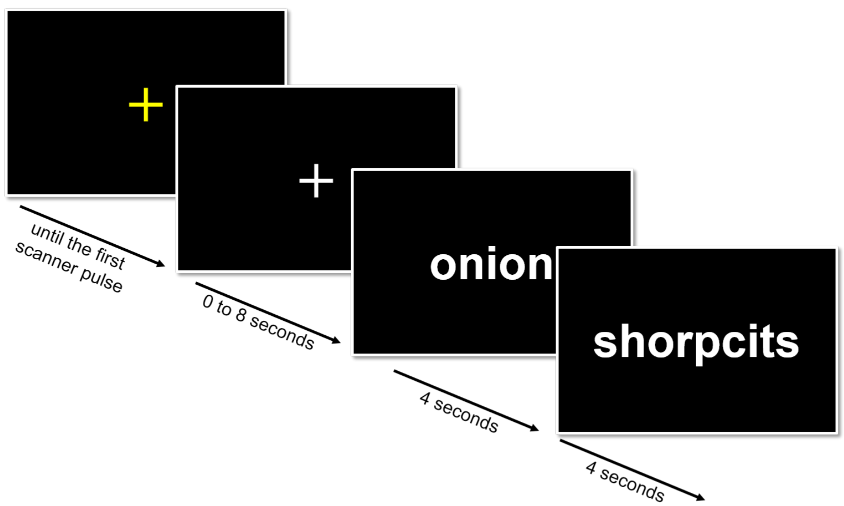

Figure 2.1. Illustration of the presentation sequence of encoding phase. Each block started with a yellow fixation which disappeared with the first scanner pulse signaling the start of each pleasantness judgment run. For the words, participants reported whether the word was pleasant or not. Nonwords were to be given the ‘NO’ response. Stimuli were presented for 4 seconds during which the judgment was rendered. ISI consisting of a fixation cross varied from 0 to 8 seconds.

Immediately following encoding, participants were removed from the scanner and given a self-paced recognition memory test in a separate room. The test items consisted of an intermixed list of studied and new words. The studied nonwords were not tested. Participants indicated if each word was old or new, followed by a three-point (low, medium, and high) confidence rating. The recognition test consisted of a total of 600 items (400 studied and 200 novel) randomly intermixed. After every 150 trials, subjects were given a brief rest before continuing. For the recognition judgment, the participants pressed 1 to indicate the item was “old” and 3 to “new”. For confidence, the 1, 2, 3 keys corresponded to each level of confidence.

After completing the recognition test, two abbreviated verbal intelligence tests were conducted. First, the participants were given a computer-based Shipley vocabulary test (Shipley, 1940). In this test, 40 synonym questions were given during which participants indicated which of four words mostly closely matched the meaning of a target word by pressing the number keys 1 to 4, indicating the position of the chosen synonym. Following this, they were given an abbreviated version of North American Adult Reading Test (NAART) which scores the pronunciation of 35 words with uncommon pronunciations (Blair & Spreen, 1989; Uttl, 2002). After completing all the tasks, the participants were debriefed and thanked for their participation.

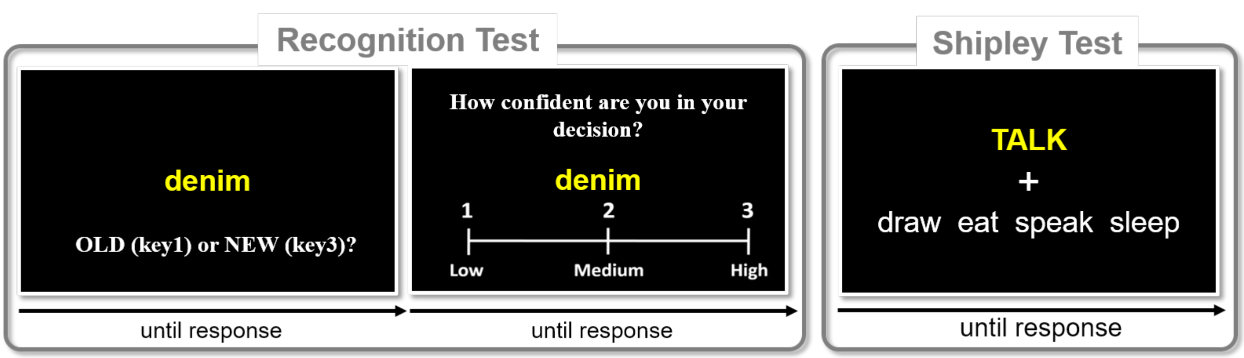

Figure 2.2. Illustration of task screen during the two computer-based tasks. During the recognition test, participants judged each word as “old” (had encountered during the encoding phase) or “new” (had never seen during the encoding phase). They also expressed their level of confidence of their judgment immediately following each recognition response. During the Shipley vocabulary test, the participants were given words among which they had to choose one word that means (nearly) the same thing as the word presented in yellow. In this example, “speak” was the right answer for the “TALK” prompt.

2.4 Design

A major feature of the study design is that every participant was presented with the same 400 words for encoding (in addition to the same 100 nonwords for encoding and an additional 200 words for recognition testing). Although the same items were presented, the order of presentation was randomized. The distribution of event types (Word, Nonword and the jittered ISIs) was determined by four sequences generated using Optseq2 algorithm. The algorithm generated fixation ISIs between 0 and 8 seconds (multiples of 2 seconds; With an ISI of 0 second, the next item immediately followed the previous item without a fixation cross). One of four sequences was then assigned to each block for each participant, whose order (which sequence for which block) was counter-balanced using a Latin-square method.

2.5 fMRI Data Acquisition

Images were acquired on a 3T Simens Magnetom Prisma fit MRI scanner (Erlangen, Germany) at the Center for Clinical Imaging Research at the Barnes-Jewish Hospital. High resolution structural data (1 x 1 x 1mm) were acquired by T1-weighted MP-RAGE sequence. The T2*-weighted functional volumes were acquired with in-plane resolution of 4 x 4 mm with 4 mm slice thickness in the FOV of 256 mm. For each volume, 34 slices were acquired in interleaved fashion from inferior to superior (TR = 2000ms, TE = 27ms, with flip angle of 90^o).

2.6 fMRI Data Analysis

Functional images were preprocessed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm) running in MATLAB R2017b (Mathworks Inc.). For motion correction, images were realigned to the first image in each session and then to the mean of all images using a rigid body spatial transformation. Slice timing onset differences were corrected using sinc-interpolation where each slice was aligned to the middle slice collected (interleaved sequence). High-resolution structural images were first co-registered to the mean functional image and were then segmented using tissue probability maps. The structural images were spatially normalized to the MNI space with forward deformation and the deformation field file created during this step was used to normalize the realigned and slice-timing-corrected functional images. The normalized functional images were then resampled to 2 x 2 x 2 mm voxels and then smoothed with 6 x 6 x 6 mm FWHM Gaussian kernel.

After preprocessing, a general linear model (GLM) was fitted for each participant's data with two conditions of interest (Words and Nonwords), various parametric modulators for "Words" trials, and six head movement parameters as regressors of no-interest. All the parametric modulation models included the reaction time for the pleasantness judgment as an additional regressor of no-interest, so that the effect of each variable could be examined without potential confounding of reaction times. To recover the unique contribution of parametric modulators, the default setting of serial orthogonalization of modulators was disabled, such that each modulator would reflect the unique contribution of that variable to predicting activation as in standard multiple regression analysis (for details, see Mumford, Poline, & Poldrack, 2015). All the parametric modulators (including reaction times) were standardized (thus, mean-centered) so that the estimate of an unmodulated regressor ("Words") represents the mean activation for the condition and the estimates of modulators reflect the parametric effect around that mean.

Chapter 3: Results

3.1 Behavioral Results

3.1.1 Recognition performance: Participant-level analysis

Table 3.1. summarizes participants' recognition performance. As expected for the deep level of processing, the average *d*-prime was relatively high. A paired *t*-test revealed that there was no difference in subsequent hit rates between words judged as pleasant versus not pleasant during encoding, $t(21) = .08$, $p = .94$. This indicates that the participants did not demonstrate a congruity effect (Craik & Tulving, 1975; Roediger & Gallo, 2001).

Table 3.1. Summary statistics for performance in the subsequent recognition memory test.

	d-prime	Bias	Overall Hits	Pleasant Hits	Not pleasant Hits	False Alarms
Mean	1.96	-.15	.85	.85	.85	.23
SD	.73	.32	.11	.13	.14	.14

With regard to the reaction times during the encoding task, when judging the words for their pleasantness, participants spent longer for the items that were later recognized ($M_{\text{Rating RT}} = 1.30$, $SD = .18$) versus forgotten ($M_{\text{Rating RT}} = 1.25$, $SD = .22$), $t(21) = 3.12$, $p = .005$. This may reflect greater elaboration at the item level which then leads to the superior subsequent memory.

3.1.2 Effect of normative item characteristics on recognition: Trial-level analysis

The goal of this analysis was to determine which of the seven normative word characteristics reliably predicted subsequent recognition memory. For the measure of subsequent recognition, instead of using dichotomized recognition outcomes (hits versus misses), I considered a more continuous measure combining outcomes with confidence ratings. This measure, named “subsequent recognition strength”, was coded with a scale of 1 through 6 spanning the highest confidence misses ‘1’ through the highest confidence hits ‘6’. Using a (more) continuous measure rather than a dichotomized one was expected to increase the power of the following analyses (Cohen, 1983).

To predict subsequent recognition strength, each normative characteristic was first considered separately in a linear mixed effect (LME) model. For the models, subjects were treated as a random effect whose intercept reflecting individual variation in mean recognition strength. Additionally, as mentioned in the Introduction, the items were also treated as a random effect, with each word modelled as having a separate intercept (Baayen et al., 2008; Freeman, Heathcote, Chalmers, & Hockley, 2010). This reflects the assumption that the words can differ in their average memorability. As I explain in section 3.2.4, this model term is a type of consensuality measure since it reflects a memorability effect of each item, collapsed across participants. However, unlike the typical memorability calculation (the proportion of subjects who correctly recognized the item), this one controls for other effects that may also be present in the data, such as variation in the average strength rating of subjects (viz., the subject intercept). All LME analyses were performed with the lme4 package (Bates et al., 2015) in the R programming language (R Core Team, 2017). As an example, the model formula examining imageability is shown below.

Recognition Strength ~ IMG + (1 | Subject) + (1 | Item)

Across the LME models, imageability, phonological distinctiveness and semantic distinctiveness yielded reliable predictions of recognition strength (All *ps* < .05). The positive estimates of imageability and semantic distinctiveness suggest that more easily imageable and semantically distinctive words were more likely to be recognized in the subsequent memory test with higher confidence rating. On the contrary, phonological distinctiveness demonstrated a negative relationship with recognition strength rating. The words that are phonologically more similar to their 20 closest neighbors, that is, less idiosyncratic in terms of their pronunciation were better recognized in subsequent testing with higher confidence.

These three variables were retained for the fMRI analyses as potential parametric modulators of encoding activation. Additionally, despite its non-significance, inverse log word frequency was also retained for a separate, subsidiary analysis, for the sake of testing its potential confounding with semantic distinctiveness which has been debated in the psycholinguistic literature (Adelman, Brown, & Quesada, 2006; Johns, Gruenenfelder, Pisoni, & Jones, 2012). Table 3.2. summarizes the estimates of the seven models (one model in each row).

Table 3.2. Estimates of each normative variable within each LME model predicting recognition strength.

Variable in each model	**Fixed effect *β***	***Std. Beta***	***t-value***	***p-value***
IMG	**.08**	**.07**	**5.22**	**< .0001**
AoA	-.04	-.02	-1.66	.098
Length	-.02	-.02	-1.43	.153
InvLogFreq	.18	.01	1.03	.305
Ortho Dist	.02	.01	.53	.594
Phono Dist	**-.09**	**-.03**	**-2.33**	**.020**
Semantic Dist	**20.32**	**.04**	**2.95**	**.003**

Finally, to ensure that the three retained variables uniquely contribute to subsequent recognition strength, I entered all three into one LME model (as shown below), again modeling both random intercepts of subject and items. The output of this model was also used during the functional imaging analysis to capture the joint contribution of the normative characteristics to encoding activations.

Recognition Strength ~ IMG + Phono Dist + Semantic Dist + (1 | Subject) + (1 | Item)

As Table 3.3 demonstrates, the variables each made a unique, reliable contribution (although it was marginally significant for phonological distinctiveness, $p = .067$) to predicting recognition strength. Although the prediction was reliable, the variance accounted for by these fixed effect variables was quite small (Marginal R^2 of .007). Together, the two random effect terms and the three fixed effect variables accounted for about 12% of the variance within the trial-wise recognition strength (Conditional R^2)[2].

Table 3.3. LME model summary with all three significant predictors of recognition strength.

	Recognition Strength			
	Beta	***Std.Beta***	***t-value***	***p-value***
Fixed Parts				
(Intercept)	-13.07		-1.96	.051
IMG	.07	0.06	4.71	< .001
Semantic Dist	18.21	0.04	2.71	.007
Phono Dist	-0.07	-0.03	-1.84	.067

[2] Marginal R^2 indicates the proportion of the total variance explained by the fixed effects whereas conditional R^2 indicates the proportion of the variance explained by both fixed and random effect (Nakagawa, Johnson, & Schielzeth, 2017)

Random Parts	
σ^2	1.678
$\tau_{00,\ \text{Word}}$	0.061
$\tau_{00,\ \text{Subject}}$	0.160
ICC_{Word}	0.032
ICC_{Subject}	0.084
Observations	8605
Marginal R^2 / Conditional R^2	0.007 / .122

3.1.3 Relationship between normative characteristics and item memorability: Item-level analysis

As mentioned earlier, unlike the trial-level analysis where the performance of individual participant in each trial was the unit of analysis, the item-level analysis collapses the performance of all the participants for an item into a single measure of consensus (viz. item memorability). Thus, the unit of analysis now becomes the item, not the trial. The goal of the analysis is to examine the relationship between two different item variables (each of the linguistic characteristics and the item memorability score), investigating the factors that potentially make some items more memorable than others at the aggregate level.

Table 3.4 summarizes the simple correlations between each normative item variable and memorability scores for the 400 encoding items. Memorability scores were either based on the performance of the current participants ($N = 22$) or taken from Cortese et al. (2015) ($N = 60$). The reliabilities of the normative variables differed somewhat when using the current sample versus the Cortes et al. (2015) memorability estimates. When combined in a multiple regression, the variables jointly accounted for 27.2% of the variance within the Cortese memorability scores (which is remarkable given the 400 nouns in this analysis are only a subset of 2,897 words from the original Cortese list), whereas they accounted for only 6.4% of the variance in the current item memorability scores.

Table 3.4. Item-level correlation between each normative variable and memorability scores from two studies.

	Memorability (Current study)	**Memorability (Cortese et al., 2015)**
IMG	**.22*** **	**.19*** **
AoA	-.08	**.18*** **
Length	-.04	-.09
InvLogFreq	.04	**.33*** **
Ortho Dist	.03	.08
Phono Dist	-.09	-.01
Semantic Dist	**.13** **	**.43*** **

Moreover, when directly compared, the two memorability scores were only modestly correlated with each other (r = .38, p < .0001), which is why the studies necessarily differ in terms of which normative characteristics are most predictive of memorability and how much variance in memorability they jointly account for. This modest correlation between the two memorability scores suggests that item memorability might be neither "normative" nor "intrinsic to the item" as discussed in the Bainbridge et al. (2013; 2017). In other words, the rank ordering of memorability of items normed from one study cannot be generalized to another without consideration of design and processing differences. In the discussion, I consider several differences between the two studies that may have produced the rank ordering differences between the studies.

It is important to note that the much smaller (compared to Cortese et al. 2015) variance in memorability accounted for by the normative characteristics at the item-level does not necessarily mean that there were smaller item memorability effects present in the current study across participants at the trial-level. In fact, it suggests that the rank order of items in terms of their memorability cannot be fully accounted for by the linguistic and semantic factors that are currently known to us. Although trial-level and item-level analyses above have demonstrated the relationship between the normative item characteristics and recognition performance, as distinguished earlier, they do not directly address the question of how much of individual's recognition outcomes can be explained by the responses that a group of subjects made to the same items. Next, I will examine the magnitude of this item memorability effects per se, moving away from any consideration of the normative item characteristics.

3.1.4 Item memorability effect within a group predicting individual's trial-wise recognition outcomes

If, on average, a large proportion of each participant's trial-wise responses can be anticipated by the response tendencies of others to the same items, we can conclude that the item effects in recognition behavior are large. Such analyses are only feasible in realistic sample sizes if all the subjects receive the same words.

To estimate the size of item effects a leave-one-out procedure was adopted in which each participant is removed from the group, and his or her recognition decisions (1='old' or 0='new') are then correlated with the proportions of remaining participants in the group who correctly recognized each item. The correlation is then saved, the participant returned, and another participant is removed, repeating the procedure. This continues until the sample is exhausted leaving N correlation coefficients reflecting the correspondence between each participant's responses and the tendencies of the remaining participants, across the studied items. Critically, if there were no item effects in the data, then this analysis would fail. Moreover, because each participant's own performance is not reflected in the memorability estimates applied to his or her prediction, the leave-one-out predictors are statistically independent of the participant's responses.

The 22 correlation coefficients calculated this way were subjected to a one sample t-test following Fisher's z transformation, demonstrating a reliable item effect in the current sample [mean Pearson $r = .11$ ($SD = .05$), $t(21) = 10.95$, $p < .0001$]. Although the absolute size of the correlation was quite small (Mean $r = .11$), we can say that item effects are nonetheless a robust phenomenon because, (a) the Cohen's *d* for the one sample t-test is 2.33 (which is "huge" according to the rules of thumb for effect size; Sawilowsky, 2009), which reflects the fact that most of the sample demonstrates a positive and similarly sized effect. In fact, the responses of 13 out of 22 participants (59.1%) were reliably predicted by the group tendencies (All *ps* $< .05$). (b) the recorded correlations are point-biserial correlations which are downwardly biased for participants whose recognition performance are either extremely good or bad. For example, a participant who correctly recognizes all of the items cannot demonstrate a correlation with the proportional response tendencies of the remaining participants because his or her responses have no variability. These restriction of range problems may be partially offset by using each subject's strength ratings instead of dichotomous outcomes. However, this would preclude comparison to the Cortese et al. (2015) data (described below) whose confidence ratings, necessary for 'strength' calculation, were not collected. Overall, the data demonstrate that a fairly small, but reliable proportion of each subject's responses can be forecasted by knowing the response tendencies of the remaining 21 subjects in the sample.

I next consider whether the size of item memorability effects in the current sample are comparable to those of Cortese et al. (2015), using the same leave-one-out procedure. A one sample t-test again revealed that the coefficients were significantly different from zero, [Mean $r = .16$, ($SD = .06$), $t(119) = 29.39$, $p < .001$][3]. Again, the Cohen's d for the one sample t-test was huge (2.68) and remarkably, the responses of 97.5% of the participants were reliably predicted by the group tendencies (All ps < .05).

Even though the leave-one-out memorability calculated from Cortese data was based on much larger sample size ($N = 59$, almost three times larger than the current sample), the mean point-biserial correlation in this group was not drastically different from that of current data. A direct comparison between the two studies indicated that the item memorability effects were larger in the Cortese study [two sample t-test: $t(140) = 3.24$, $p < .005$] but the difference between the mean correlation coefficients was quite small ($r = .04$). This suggests that the smaller sample size in the current study may have led to memorability estimates that are somewhat noisier (i.e., more sensitive to each subject's removal) than in Cortese et al. However, this relatively small difference between the two studies, compared to the drastic difference shown in the normative

[3] In Cortese et al. (2015), each word was presented as both target and lure to a random set of 60 different participants in a group of 120. Given this assignment protocol, no two participants had the same target and lure lists, so I could not restrict the analysis to the 400 nouns selected for the current study. Thus, for Cortese data, the leave-one-out memorability scores are based on the 59 (out of 60) participants who were given the same word as a target. To clarify, in Cortese et al., each of 120 participants received 1,500 target words, but each word was given to a set of 60 random participants. Thus, the leave-one out procedure was performed for a total of 120 participants (hence the DF for one sample t-test was 119, not 59) but the memorability calculated for each word was based on 59 participants, not 119.

characteristics approach (section 3.1.2 and 3.1.4) supports the distinction between the item memorability approach and item characteristics approach introduced earlier.

Overall, these results suggest that just by knowing the performance of others within a sample, one can reliably predict a portion of each individual's responses (*rs* > .11) and do so for the majority of the participants in the sample (> 59%). This demonstrates that there are item effects based on the item memorability which presumably rely upon the fact people share common word representations.

3.1.5 Linking item effects to individual differences in verbal IQ

If item effects reflect the contribution of linguistic representations to encoding efficacy, then the effects are potentially limited by the observer's familiarity and understanding of verbal items (presumably measured via verbal intelligence). For example, a semantically distinctive item cannot be encoded as such for a subject who does not know its definition.

This possibility can be examined by first correlating each participant's recognition strength ratings with each of the three normative variables (IMG, Semantic Dist and Phono Dist) that were predictive at the trial-level analysis.[4] Thus, for each participant, I obtained three correlation coefficients tracking the degree to which his or her recognition strength was

---

[4] Note that the random by-subject slope for each variable in LME model can be also used to address the same relationship (Barr et al., 2013). However, among the variable models introduced in the section 3.1.2, only a subset converged with the random slope term. Thus, to keep it consistent for every variable, I chose the LME models without random slope as the variable selection criteria. This summary approach for the simple correlations is then, another way of expressing that the slope/strength of relationship between a recognition response and each variable can differ across participants.

influenced by each of the three normative item characteristics. The three coefficients were then correlated with each subject's Shipley and NAART35 scores to see whether the influence of the normative item characteristics was linked to these short measures of verbal IQ. Before presenting that analysis I briefly summarize the group performance on these IQ measures below.

Overall, the Shipley test yielded an average proportion correct of .82 (*SD* = .09) and the NAART35 yielded a slightly lower score of .72 (*SD* = .12). One participant with reasonable recognition performance (*proportion correct* = .69 and *d-prime* = .84) nonetheless demonstrated exceptionally low performance for both Verbal IQ tests (.5 for Shipley and .49 for NAART35; which is 3 *SD*s below the mean for Shipley, and 1.9 for NAART35). Because the person's recognition performance did not disqualify him/her for the overall analyses (both behavioral and fMRI), below I report the Verbal IQ correlation results with and without the participant's data.

Table 3.5 shows the degree to which the three normative variable-to-recognition strength correlations for each subject are correlated with their performance on the Shipley vocabulary tests. Of the three variables, only the correlation between semantic distinctiveness and recognition strength was reliably associated with the observers' Shipley vocabulary scores, both with and without the potential outlier. This suggests that people with better semantic/vocabulary knowledge obtain more benefit from variations in semantic distinctiveness of items either during encoding and/or retrieval. In other words, the degree to which an individual better recognizes semantically distinctive items than less distinctive items depends upon his or her depth of semantic knowledge.

Table 3.5. Correlation between individual's Shipley score and variable-to-recognition correlation

Influence on recognition from	**With Verbal IQ outlier**		**Without Verbal IQ outlier**	
	Pearson r	***p-value***	***Pearson r***	***p-value***

IMG	-.03	.18	-.27	.24
Phono Dist	.10	.66	-.03	.90
Semantic Dist	.68	< .001	.47	< .05

The procedure was repeated for NAART35 scores, and the findings are summarized in table 3.6. As with Shipley scores, imageability and phonological distinctiveness effects were again unrelated to individual difference in the NAART35 performance. Partially converging on the Shipley findings, semantic distinctiveness effects were reliably correlated with NAART35 performance. However, this was only reliable when the potential outlier was included in the analysis.

Table 3.6. Correlation between individual's NAART 35 score and variable-to-recognition correlation

Influence on recognition from	**With Verbal IQ outlier**		**Without Verbal IQ outlier**	
	Pearson r	***p-value***	***Pearson r***	***p-value***
IMG	-.04	.85	.03	.90
Phono Dist	-.10	.67	-.18	.42
Semantic Dist	.42	.05	.24	.29

Overall, the data provide some evidence that verbal IQ moderates the influence of semantic distinctiveness of items on encoding (although care should be exercised given the modest sample size for individual differences analyses). Critically, neither Shipley nor NAART35 scores demonstrated a significant correlation with participants' recognition *d-prime* (all *ps* > .42), suggesting that the participants' Verbal IQ had no direct relationship to their recognition memory for verbal items. (Note that there are studies reporting a reliable relationship between fluid intelligence and recall or associative recognition (Healey, Crutchley, & Kahana, 2014; Ratcliff, Thapar, & McKoon, 2011). Thus, current findings suggest that the individuals with better verbal IQ do not necessarily have better recognition memory, but they do get more mnemonic benefits from semantic distinctiveness of items presumably due to their sensitivity to this information.

3.2 fMRI Analysis Results

The plan of the fMRI data analyses largely parallel those of the behavioral analyses. I begin by documenting the basic subsequent memory effect without the consideration of item effects, to ensure it replicates prior findings. Analyses concerning item effects will follow, to examine the degree to which they account for the basic subsequent memory activation.

All the activation maps shown below were based on the voxel-wise comparisons thresholded at the level of $p < .001$ (uncorrected) significant for minimum of 5 contiguous voxels. Full SPM tables of suprathreshold regions in each activation map (positive effects only) are listed in the Appendix II.

3.2.1 The basic subsequent memory effect

(1) Regions linked to subsequent recognition memory

As in behavioral analyses, the recognition outcome was re-coded as recognition strength (1 = high confidence miss to 6 = high confidence hit) so that it suits later LME modelling framework. The strength value was entered as a parametric modulator to identify the regions whose activation during encoding, forecast the level of subsequent recognition. The regressor yielded reliable positive activations in bilateral PMC/supplementary motor area (SMA) [BA 6, 8], left IFC [BA 44, 47], bilateral superior parietal lobule (SPL) [BA 7], bilateral angular gyrus

(AnG) [BA 39], and bilateral occipital regions [BA 19]. Subsequent recognition strength also activated bilateral ventral temporal regions including fusiform gyrus (FuG) [BA 20, 37], potentially extending into the hippocampal formation.

Although the continuous, recognition strength reports were the main regressor throughout the analyses, subsequent memory research often focuses on the dichotomous contrast of hits versus misses. As Figure 3.1 shows, the two variables yield highly overlapping maps. In the bottom row, the contrast map is overlaid with the recognition strength map (indicated by the white outline) for visual comparison. More concretely, 78.95 % of the voxels of contrast map fell within the strength boundary showing considerable overlap between the two activation maps. Thus, hereafter I will refer to the recognition strength map when discussing the subsequent memory effect.

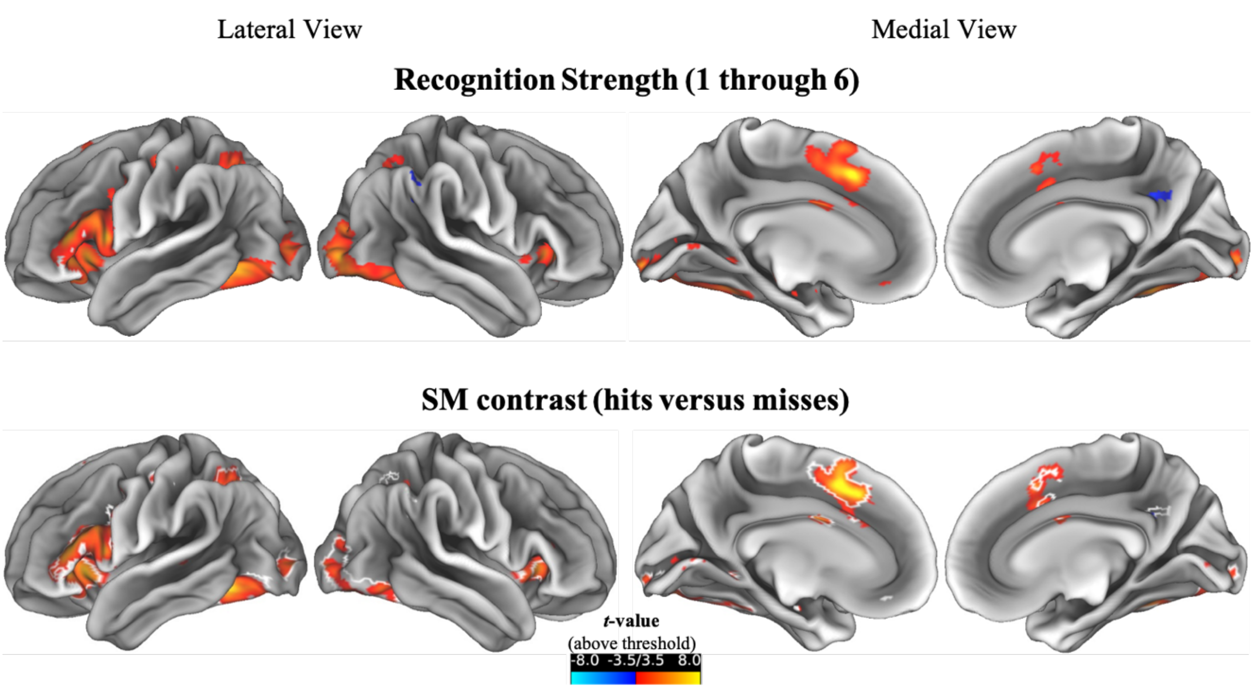

Figure 3.1. Subsequent memory effect maps defined by continuous recognition strength variable (upper panel) and traditional dichotomous contrast (lower panel). In upper, the warmer color indicates regions showing greater activity for increasing level of recognition strength (the positive effect) and the cooler

color indicates regions showing the opposite effect. In lower panel, the warmer color represents regions showing greater activation for subsequently recognized words (hits) than forgotten words (misses), vice versa for the cooler color. White outline in the lower panel represents the outlines of regions shown in the upper panel, for visual comparison between the two maps.

(2) Comparison to meta-analytic SME map

The current subsequent memory effects demonstrated overlap with SMEs reported in the literature as illustrated by the meta-analysis conducted by Kim (2011). Figure 3.2 shows the current recognition strength map (in blue) overlaid with the meta-analysis subsequent memory map (in red). There are several areas of overlap including left IFC, SMA/middle cingulate and infero-temporal regions. However, it is also clear that the current effects include several areas that are not in the meta-analytic map such as bilateral inferior frontal regions including anterior insula (aIns)/ posterior orbital gyrus (pOrG), and inferior posterior regions including extrastriate and more extensive fusiform areas. These additional activations may have resulted from restricting the materials to nouns, insuring through selection that they had a good range of item memorability, or from administering the same set of items to all participants.

Strength (blue) overlaid with Meta-analysis (Kim, 2011) SME map for *verbal item* (red)

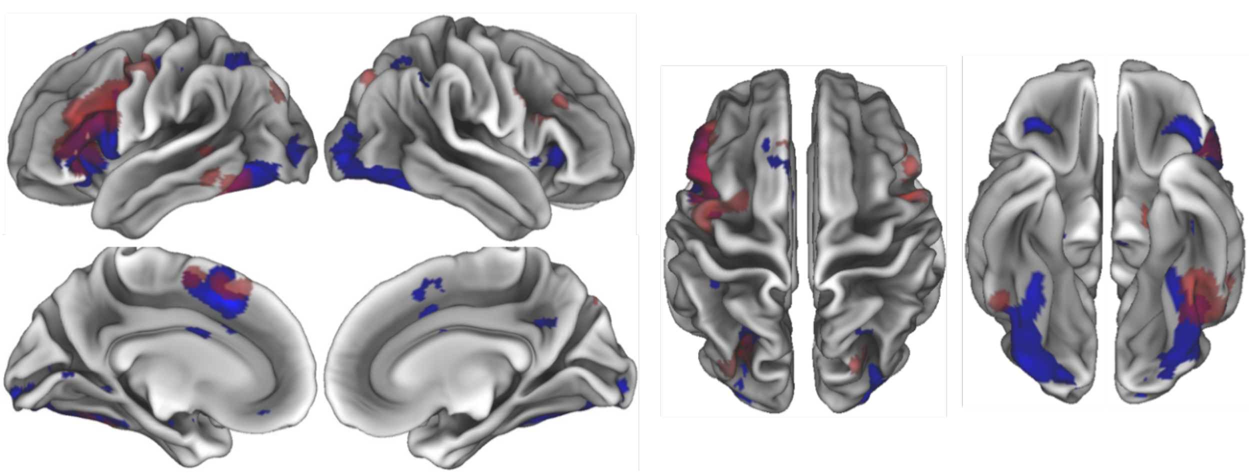

Figure 3.2. Regions shown in blue are the parametric modulation map of recognition strength and the regions shown in red are the subsequent memory map from a meta-analysis (Kim, 2011). The meta-analysis map is based on the studies that used verbal materials as their stimuli and tested for item memory (as opposed to associative memory) to match to the current stimuli and procedure.

3.2.2 Parametric modulation via normative item characteristics

As noted in the behavioral analyses, three normative item characteristics (imageability, phonological and semantic distinctiveness) predicted recognition strength ratings (Table 3.2). Item effects in the SME activation will be examined first focusing on each in isolation through parametric regression of each variable. Figure 3.3 illustrates the effects of each normative characteristic during encoding. As a reminder, reaction time for pleasantness judgment was included as a covariate of no interest in every analysis, properly partialling out the potential confounding of (incidental) encoding duration. Thus, these activations reflect responses not attributable to slower judgement. Each parametric modulation map is overlaid with the boundary of the subsequent recognition strength map (white outline) for visual comparison.

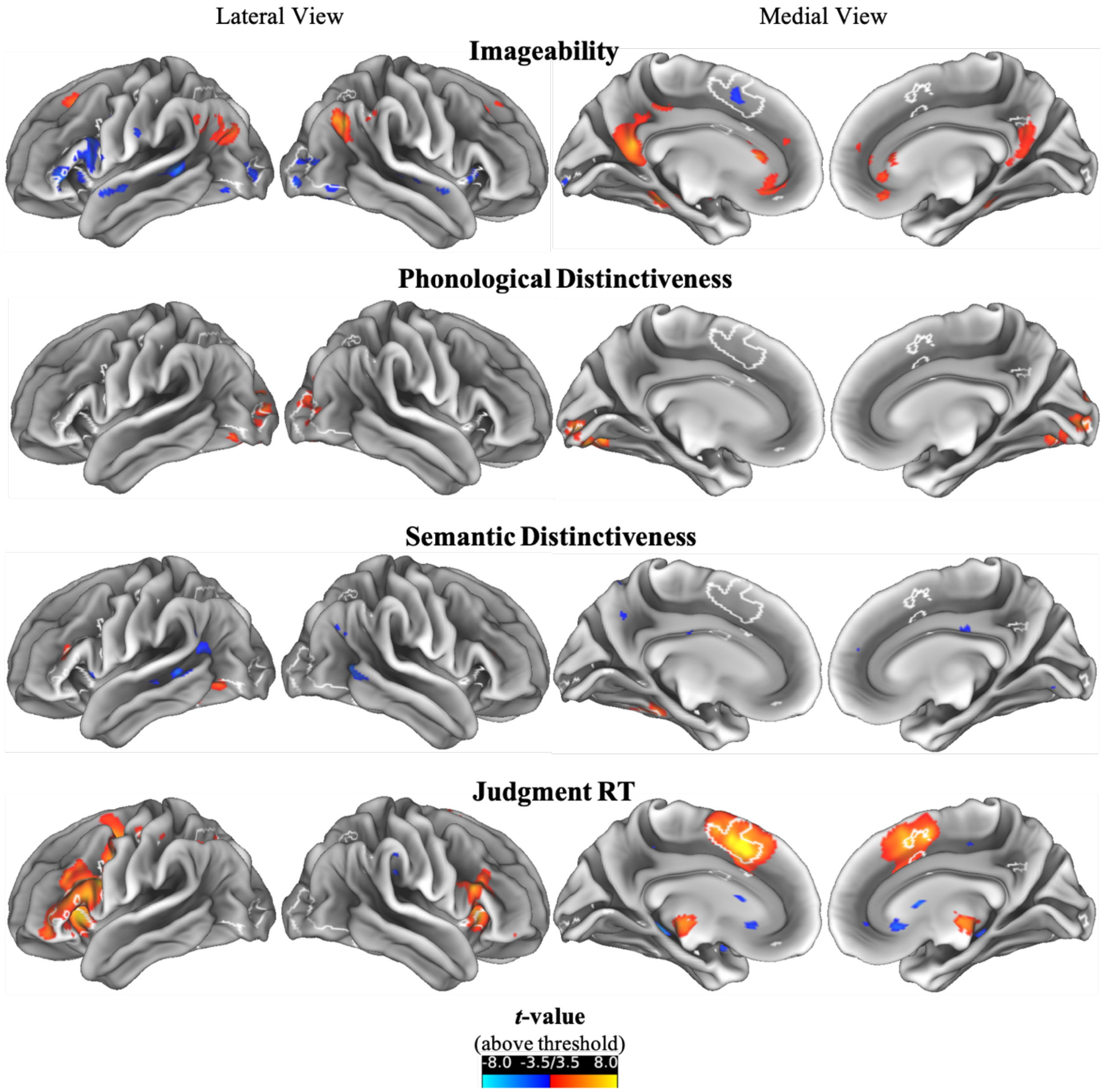

Figure 3.3. Parametric modulation of normative item characteristics. The warmer color represents the regions that showed greater activation for increasing value of each variable (positive effect) whereas the cooler color represents the regions that showed greater activation for decreasing value of each variable (negative effect). The white outlines indicate the boundary of the regions showing the subsequent recognition strength effect.

Table 3.7 summarizes the pattern of overlap between each parametric modulation map and the subsequent recognition strength map. The percent overlap was calculated from the number of voxels falling within the subsequent recognition strength map that are significantly activated by the normative characteristic, divided by the number of total voxels activated by the normative characteristics.

Table 3.7. Overlap between the normative characteristic maps and the recognition strength map.

Item Variables*	# of total voxels	# of voxels fall within the recognition strength map	% Overlap with recognition strength map
IMG	4019	32	0.80%
Phono Dist	3012	932	30.94%
Semantic Dist	738	474	64.23%

As shown in Figure 3.3, increasing imageability produced activation in dorsal prefrontal and lateral parietal areas, along with posterior cingulate and retrosplenial areas. However, despite the largest (among the three variables) contribution of imageability in the prediction of recognition strength, the positive activation map of imageability did not overlap much with the recognition strength map. In fact, in terms of the number of voxels that fell within the recognition strength map (Table 3.7), imageability was the variable that showed the least overlap (only .8%) with the recognition strength map.

Phonological distinctiveness showed activation that was quite restricted to the occipital lobe. This may suggest that the phonologically more distinctive words required more visual processing. The high positive correlation between phonological distinctiveness and orthographic distinctiveness (r = .61) suggests that these words might have required more intensive visual processing, which again explains the localized activation in the occipital area. In fact, orthographic distinctiveness (not shown here) produced activation almost identical to phonological distinctiveness map without much unique activation of either when pitted against each other. In terms of overlap with recognition strength map, although about 30% of the voxels showing positive effect fell within the recognition strength boundary, this must be interpreted with caution given that behaviorally, phonological distinctiveness had a small but negative (not positive) effect on recognition strength. Critically, despite the negative relationship between phonological distinctiveness and subsequent recognition behavior, there were no suprathreshold clusters that tracked a decreasing degree of phonological distinctiveness.

The semantic distinctiveness produced significant activation in bilateral (but mostly left-lateralized) IFC and fusiform/parahippocampal area (regions hereafter denoted as ventral MTL). Because of the relatively smaller area of activation which fell mostly within the recognition strength map, among the three variables, semantic distinctiveness showed the greatest overlap with the recognition strength activation in terms of voxel count (Table 3.7).

Finally, as mentioned earlier, although I regressed out the pleasantness judgment reaction time of each participant when modeling parametric modulation of the normative item variables as well as when creating the recognition strength map, Figure 3.3 also includes the parametric modulation of the reaction time itself producing a robust activation in bilateral ventrolateral and dorsomedial prefrontal areas. Thus, the recognition strength map shown in Figure 3.1 is the subsequent recognition map after these reaction time influences have been partialled out.

3.2.3 Unique modulations of normative item characteristics

(1) Three item characteristics that predict subsequent recognition strength

Behaviorally, the multiple regression model in Table 3.3 demonstrated that three normative item characteristics made unique contributions to recognition strength reports. Here, I conduct an fMRI analysis analogous to the multiple regression, in order to identify activation during encoding that is uniquely tied to each characteristic by modelling all three characteristics simultaneously as parametric modulators. Thus, in this framework, the reliable activations linked to any one characteristic represent modulatory effects that are unique in the presence of the other two.

As shown in Figure 3.4, imageability implicated multiple fronto-parietal regions including bilateral AnG [50, -4, 36; -42, -76, 32], posterior cingulate/precuneus (PCgG/PCu) [-4, -54, 18; 8, -54, 14] and ventral MTL [-32, -34, -14; 32, -32, -18]. The fronto-parietal responses may reflect visual processing/inspection processes recruited for highly imageable/concrete items in addition to the PCgG/PCu response often linked to visual imagery (Cavanna & Trimble, 2006; Fletcher et al., 1995; Ganis, Thompson, & Kosslyn, 2004). The top three scoring items on this attribute were ‘blizzard’, ‘bullet’ and ‘cigar’. In contrast, phonological distinctiveness (blue) was linked to extrastriate responses [-10, -82, -14; 20, -94, 12] and likely reflects the fact that phonologically distinctive items also tend to have unusual word forms. The top three items for phonological distinctiveness were ‘platform’, ‘pigsty’ and ‘penguin’. (Note that these items were less likely to be subsequently recognized.). Finally, semantic distinctiveness activated left IFG [-44, 28, 10] and left ventral MTL [-32 -44, -18], both of which fell within the main subsequent recognition strength map. The top three items for semantic distinctiveness were ‘tassel’, ‘talcum’ and ‘rumba’. Interestingly, Figure 3.4 demonstrates a very proximal region responds to items that are increasingly imageable during encoding and suggests that left ventral MTL may be a convergence zone for various types of representational distinctiveness; a possibility I consider next using LME modeling estimates.

Critically, this map illustrates the unique influence of these three normative characteristics which also behaviorally predict recognition strength outcomes. Nonetheless, because they predominantly fall outside the main subsequent recognition strength map, this provides the first piece of evidence that the SME map might largely reflect subject-driven and not item-driven encoding phenomena.

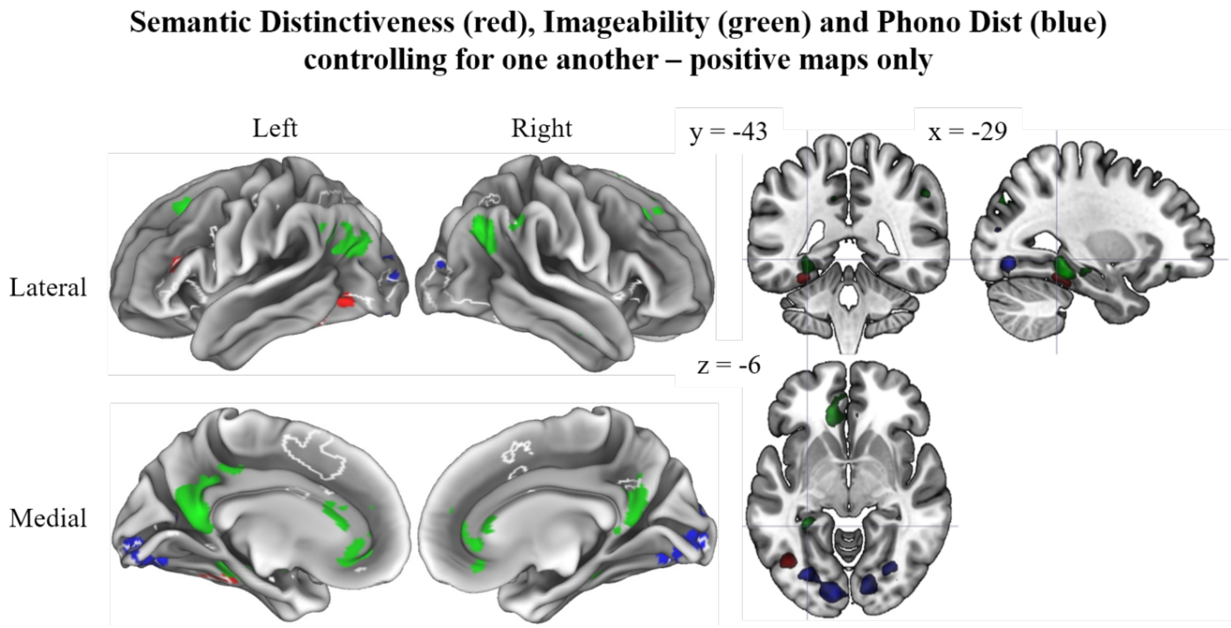

Figure 3.4. Parametric modulation map demonstrating the unique influence of three item characteristics, overlaid with outline of recognition strength map (white).

(2) Semantic distinctiveness versus word frequency

Although word frequency is an important predictor for episodic memory (Balota, Burgess, Cortese, & Adams, 2002; Balota & Neely, 1980; Criss & Malmberg, 2008; Glanzer & Adams, 1985; Hemmer & Criss, 2013; Malmberg & Nelson, 2003; Park, Reder, & Dickison, 2005), its potential confounding with more semantic measures such as contextual diversity or semantic distinctiveness has been a source of debate mostly within the psycholinguistic literature (Adelman et al., 2006; Johns et al., 2012), not within the episodic memory literature. In fact, there have been very few episodic memory studies that treated word frequency as a continuous variable (Cortese et al., 2015; Hemmer & Criss, 2013; Lau et al., 2018). The current study provides a good opportunity to study the issue of potential interaction or dissociation between the two variables in memory.

For the 400 encoding items in current study, (inverse) word frequency showed a fairly large positive correlation with semantic distinctiveness (r = .61). Despite their high correlation, in a study comparing relative contribution of lexical versus semantic variables, Lau and colleagues (2018) suggested that the lexical variables including frequency jointly accounted for more variance in recognition than in recall, whereas semantic variables explained additional variance in recall on the top of what was explained by lexical variables. Together, these raise a possibility that the two variables capture similar but distinguishable psychological processes during encoding.

Given the relationship between frequency and semantic distinctiveness, I tested whether the regions displaying semantic distinctiveness effects during encoding (Figure 3.3) remained active when pitted directly against inverse log word frequency. As Figure 3.5 demonstrates, in the presence of the inverse word frequency, left FuG/PHG [-28, -34, -18] remained linked to semantic distinctiveness although the peak of the cluster moved anteriorly, so that the major portion of the cluster is no longer within the recognition strength boundaries. Also, the left IFG response previously linked to semantic distinctiveness was no longer observed.

On the other hand, inverse word frequency itself produced significant activations in several areas (e.g., IFG, SMA, and widespread regions in ventro-temporal/occipital area). The ventro-temporal/occipital activation is of particular interest because it replicates the previous psycholinguistics findings demonstrating that the putative VWFA is not limited to sublexical processing of "word form" but is involved with more lexical properties of words (Schuster et al., 2016; Yarkoni, Speer, et al., 2008) or more abstract processing not restricted to words (Kronbichler et al., 2004; Price & Devlin, 2003; Vogel, Miezin, Petersen, & Schlaggar, 2012; Vogel et al., 2014) However, despite the region's theoretical importance and its overlap with recognition strength map, it is difficult to conclude that the activation contributes to recognition performance given the failure of inverse word frequency in predicting recognition strength at the behavioral level (however, see Hemmer & Criss, 2013 for potential non-linear relationship between word frequency and recognition memory).

Taken together, the findings suggest semantic distinctiveness produces left ventral MTL (aFuG/pPHG) activation that is reliable even after controlling for the influence of inverse word frequency despite the high correlation between the two variables. In addition, the fact that the semantic distinctiveness activation moved slightly forward (anteriorly) in the presence of inverse word frequency suggests that the posterior part of the aFuG/pPHG activation shown before might have been conflated with potential influence of word frequency. Similar "adjustment" is observed when the three significant behavioral predictors of recognition strength (viz., IMG, PhonoDist, SemanticDist) provide joint prediction (as opposed to unique contribution examined in the previous section 3.2.4. (1)) within an LME modeling framework, which will be introduced in the next section.

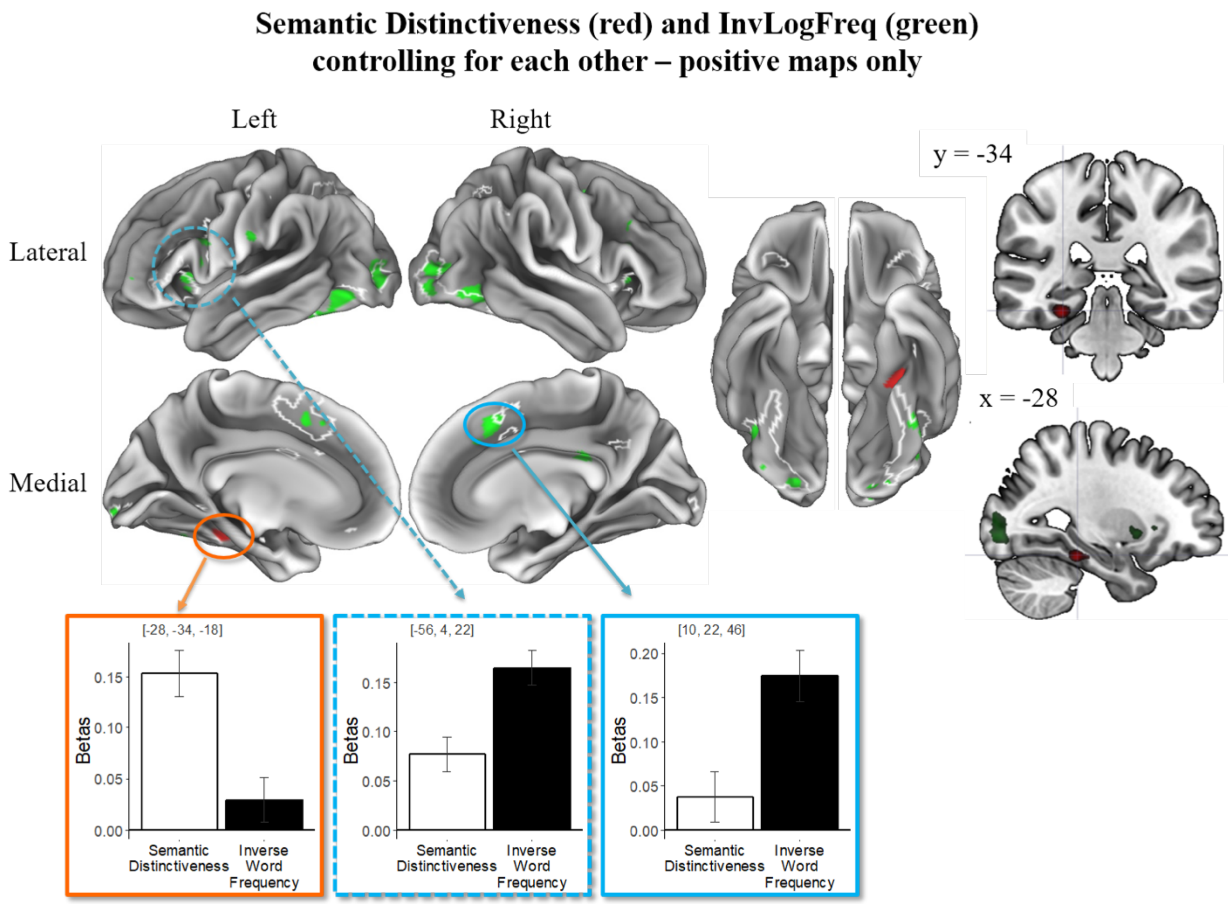

Figure 3.5. Unique parametric modulation of semantic distinctiveness and inverse word frequency controlling for each other.

3.2.4 Isolating components of encoding activation using an LME model

The above analysis demonstrates that imageability, semantic distinctiveness and (inverse) phonological distinctiveness each make **unique** contributions to activations during encoding. This is important, because each variable was shown to reliably predict participants' recognition strength reports, but presumably cannot be represented in a common manner. For example, the representational structure that makes semantics distinctive should be separable from the structure that makes phonological features distinctive, at least in early processing stream. However, Figure 3.4 demonstrates that semantic distinctiveness and imageability activate proximal regions of left ventral MTL (aFuG/pPHG). Based on models assuming parahippocampal regions as potential convergence zone for information travelling to the hippocampal formation (Aggleton & Brown, 1999), I next turned to modeling the joint (as opposed to unique) contribution of the identified normative item characteristics to subsequent recognition strength. To do this, I used the LME model fit to the behavioral strength reports of the subjects.

As noted in section 3.1.2, this model contained three reliable fixed effects of the normative characteristics discussed above, and a random intercept of items, reflecting word-level effects not covered by these three characteristics. In total, the model in Equation 3.1 provides three separate components that can be used to predict encoding activation at the level of each trial.

<u>Equation 3.1</u>

Recognition Strength ~ IMG + Semantic Dist + Phono Dist + (1 | Subject) + (1 | Item)

The first three fixed effect terms represent the normative characteristics shown to influence recognition strengths in behavior, which also provided unique activations when used as parametric regressors during encoding (Figure 3.4). I refer to the model-determined weighted sum of these three variables as ‘item distinctiveness’ based on the hypothesis that highly imageable and semantically distinctive items will yield durable memory representations. As to phonological distinctiveness, it is difficult to make a simple hypothesis because the direction of its relationship to recognition strength is opposite to the other two variables (note that phonological distinctiveness consistently demonstrated negative relationships to recognition performance measures in trial-level LME models as well as in item-level simple correlations). A simple explanation of this pattern is that, items with lower phonological distinctiveness may be easier to read, which may facilitate subsequent processing of the items (e.g., at the semantic level). However, its positive correlation to orthographic distinctiveness which, unlike phonological distinctiveness, demonstrated a positive correlation to recognition (non-significant in the current study but significant in Cortese et al. (2015)), suggests that the relationship between phonological and orthographic distinctiveness and their influence on episodic memory might be more complex than what a simple linear association can tell us. For example, distinctive mapping between orthography and phonology or having fewer phonological-to-orthographical neighbors (Cortese, Watson, Wang, & Fugett, 2004; Hirshman & Jackson, 1997) has been demonstrated to produce better episodic memory. This pattern of mapping or consistency between orthography and phonology cannot be captured by the simple correlation between the two distinctiveness measures.

The next term of the equation, random intercept of subjects captures mean differences across subjects in the rated strength of the items, capturing for example, individual differences in scale use or differences in the tendency towards caution. The final random effect term, '(1 | Item)' captures the tendency (collapsed across subjects) of each word towards a particular mean strength report level, and hence captures item effects not specifically modelled by the three normative item characteristics in the equation or subject differences in mean strength rating. Conceptually, this term reflects unknown but systematic influences on the mean rated strength across the items presumably representing normative characteristics of which I am currently unaware.

In total, this fitted model captures all currently known item effects in the recognition data and it is sufficiently powered because each subject received exactly the same set of encoding items. Once fit, the model components were then used, via the ‘predict.merMod()’ function of the lme4 package, to produce adjusted covariates linking behavioral item effects to fMRI activation at the trial-level for specific components of the model.

(1) Fixed effect prediction from three normative variables - Item Distinctiveness Component

First, I consider the joint influence of the three normative variables theorized to reflect item distinctiveness. As Figure 3.6 shows, item distinctiveness activates largely three sets of regions; first, ventral MTL regions including bilateral aFuG/pPHG [-30, -36, -18; 30, -32, -20] (BA 36/37) and some portion of left hippocampus [-34, -26, -18] (BA 54), second, bilateral inferior frontal gyrus (IFG) including aIns/pOrG [28, 30, -10; -28, 32, -10] (BA 47) and finally, left posterior cingulate/ventral precuneus [-4, -54, 16] (BA 23). As will be shown in the following section, the latter two are largely overlapping with regions activating for random intercept prediction. The ventral MTL activations which are unique to the item distinctiveness terms are also anterior to the initial subsequent recognition strength map boundaries. Critically, these regions were only discovered because item distinctiveness effects were specifically modelled. As noted in section 3.1.4 of the behavioral analysis, item-level effects account for a small, but reliable portion of each subject’s memory responding. Hence failure to model them specifically will likely cause them to be missed because they will be swamped by the portion of strength ratings not governed by item-driven effects; which in this case is the majority. Hence the LME model provides the ability to target specific influences not easily gleaned from the overall basic subsequent recognition analysis.

Aside from the location of the bilateral ventral MTL responses, and the nature of the normative characteristics driving this response, are there other reasons to conclude that the **distinctiveness interpretation** is correct? I further explore this by noting that a distinctiveness interpretation can also be tested using the consensual false alarm rates from Cortese et al. (2015). In Cortese et al., not only was each item normed for the tendency of subjects to correctly recognize it but was also normed for the tendency of subjects to false alarm to it during recognition. If the combination of the three normative variables truly reflects item distinctiveness, it should negatively correlate with these false alarm tendencies obtained from that independent data set. A reliable negative correlation between the item distinctiveness estimates and Cortese FAs across the items (r = -.26, p < .001) supported this hypothesis. This "mirror effect" of item distinctiveness term supports the idea that the relatively distinctive features not only serve as a basis for durable encoding, but that the absence of memorial information given these features is used as a basis for rejecting them as studied (Brown, Lewis, & Monk, 1977).

Predicted value from three normative variables (Item Distinctiveness)

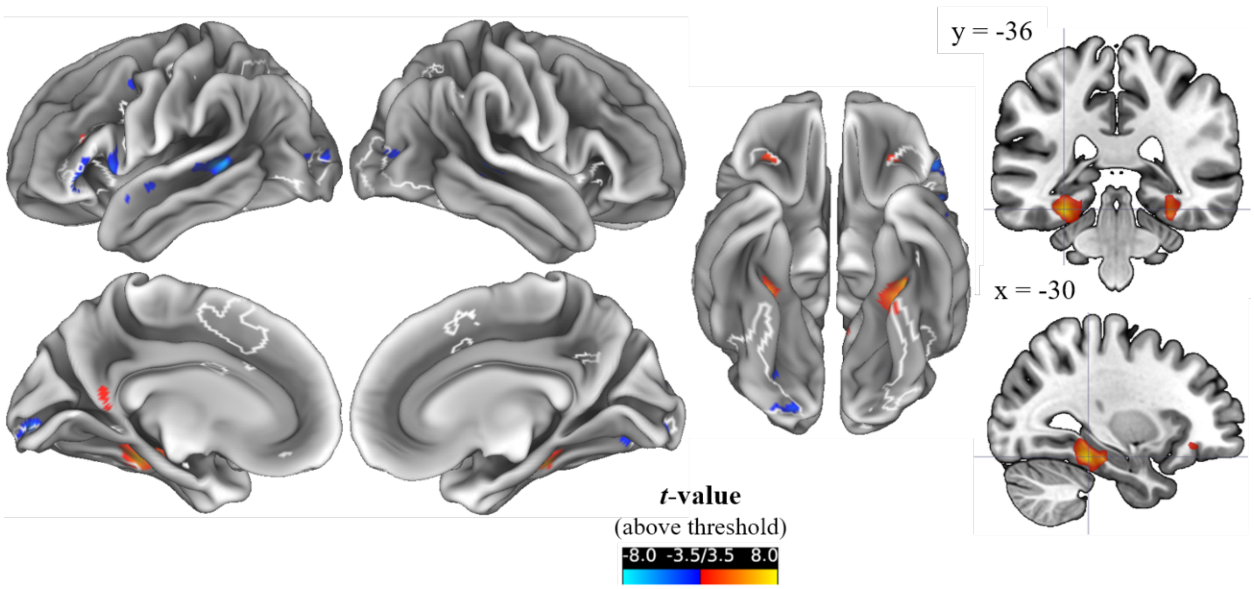

Figure 3.6. Parametric modulation of the LME model prediction based on "joint" contribution of the three fixed effect item variables (item distinctiveness component of encoding activation).

(2) Random intercepts of items – general item memorability component

As noted above, the '(1 | Item)' term in Equation 3.1 captures item effects not explicitly picked up by the three normative variables constituting item distinctiveness. I refer to this model component as 'mean item memorability'. When this portion of the model is used to generate an fMRI covariate, left dorsomedial prefrontal cortex (DMPFC) [-6, 46, 24] (BA 9/10) (which is inferior to the DMPFC activation within the recognition strength map) were identified, along with left posterior cingulate/precuneus [-8, -52, 22] (BA 23) and bilateral IFG also including aIns/pOrG [30, 30, -8; -34, 32, -8] (BA 45/47) which extends to frontal operculum on the left side [-32, 30, 6] (Figure 3.7).

The bilateral IFG responses appears to overlap with those arising from item distinctiveness terms (although here the regions are more extended) consistent with the possibility that this region signals general item salience or attentional capture; a possibility considered in the discussion. The posterior cingulate/precuneus responses also seem to be largely overlapping with the item distinctiveness component although they might be slightly superior to those shown in distinctiveness component. Overall, the overlapping regions generally showed more robust/extended responses to the random item intercept (item memorability) than to the fixed effect (item distinctiveness) prediction.

Finally, unlike other regions that are shared with fixed effect predictions, the left DMPFC responses were unique to item intercept prediction. Interpreting the role of this DMPFC regions is, however, much trickier since it is unclear whether this cluster is part of the salience network along with the IFG activations or whether it is part of the default network (Buckner, Andrews-Hanna, & Schacter, 2008; Raichle et al., 2001) together with posterior cingulate (Power et al., 2011; Yeo et al., 2011) without further analyses necessary for demarcation (e.g., functional connectivity analysis with better spatial resolution). The potential role of these regions will be suggested in discussion but given the exploratory nature of the current dissertation, strong claims regarding the functional roles of each cluster would not be suitable for the current data.

Estimated Random Intercept of Item (Mean Item Memorability)

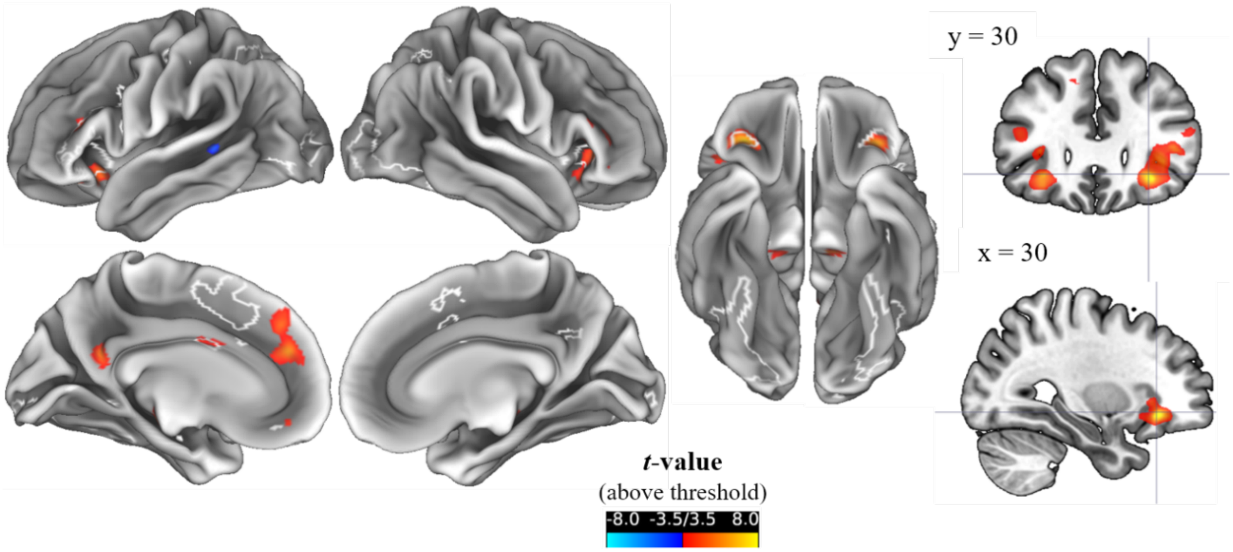

Figure 3.7. Parametric modulation of LME model prediction based on random intercept of item (mean item memorability component of encoding activation).

What does the random intercept of items capture?

As demonstrated in the behavioral analyses in section 3.1.3, the consensual item memorability estimates calculated for the current sample were modestly correlated with the estimates of Cortese et al. (2015) across the 400 encoded items (r = .38, p < .0001). As noted above, the prediction from the term '(1 | Item)' of the LME model in Equation 3.1 should also reflect item memorability (but not measured as the consensual hit rate, but as predicted recognition strengths), albeit memorability controlling for variation due to item distinctiveness and subject differences in their mean strength ratings. Given this, the model-based mean item memorability estimates should correlate with the consensual memorability estimates in the current sample. It does, yielding a reliable correlation (r = .90, p < .001) across the 400 items. Moreover, the model-based mean item memorability term modestly correlates (r = .34, p < .001) with the consensual memorability estimates from Cortese et al. (2015), replicating the modest correlation seen between the memorability scores of the two studies. Practically, this demonstrates that one can recover item-driven effects without LME modeling simply by entering the consensual hit rate for each item. However, it should be noted that this approach does not allow one to recover the item distinctiveness effect, nor does it remove subject variation in mean strength (or decision biases) from the estimates.

In terms of encoding-related activations, Figure 3.8 shows that the item intercept term and the leave-one-out memorability track activation in essentially the same regions and thus the behavioral and fMRI data demonstrate the two measures are essentially proxies for one another.

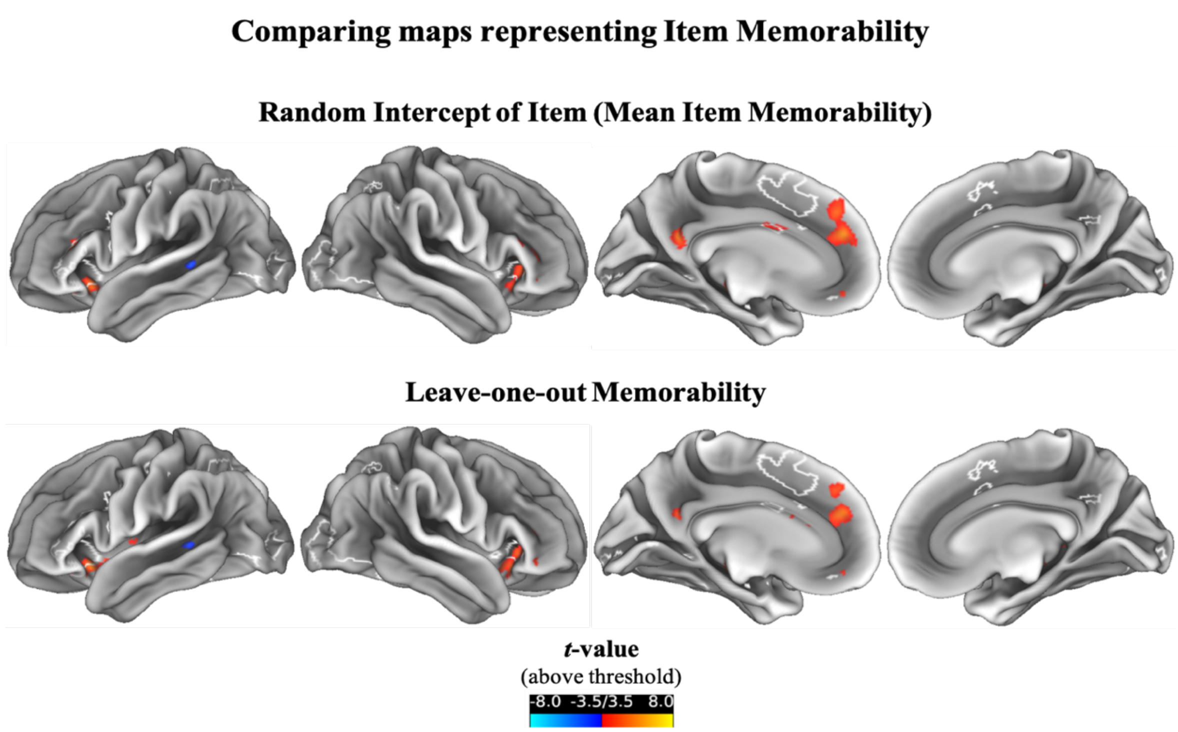

Figure 3.8. Parametric modulation of different item memorability measures.

(3) Residuals of the item effects LME model – recognition strength unexplained by items

As noted above, the LME in Equation 3.1 is the complete model of item effects in our current data encompassing the effects attributable to three known normative item characteristics as well as the effects from general memorability differences across items for unknown reasons. Consequently, the residuals of this model reflect variance in recognition strength reports that **cannot** be explained via item effects. I refer to any effects linked to these residual values as 'subject-driven' SMEs because they are not explainable as a function of item-related terms. Using the residuals to interrogate encoding activation yields the map in Figure 3.9. Critically, the model residual map not only recovered most of the original recognition strength map (which used raw strength scores as the covariate), but also demonstrated some expansion of the regions (specifically, additional 1,112 voxels were discovered outside of the recognition strength mask). Given that the residual of the model is the recognition strengths unaccounted for by item-driven effects (either mean item memorability or item distinctiveness), the expanded map can be considered the SME map after correctly controlling for potential item effects. The fact that it is more robust than the original recognition strength map (white outlines), revealing more suprathreshold regions adjacent to boundaries of the original strength map, potentially reflects the increased power gained from statistically controlling the item effects. Moreover, the fact that the residuals mostly recovered the original recognition strength map demonstrates that the effects within the original map are predominantly subject-driven; reflecting processes that are common across participants (hence, detectable via the second-level random effect (RFX) analysis in SPM), yet unpredictable by knowing the identity of the specific item the subject is processing at each trial. I discuss the psychological importance of distinguishing subject- versus item-driven encoding phenomena further in the discussion section.

Item Effects Model Residual

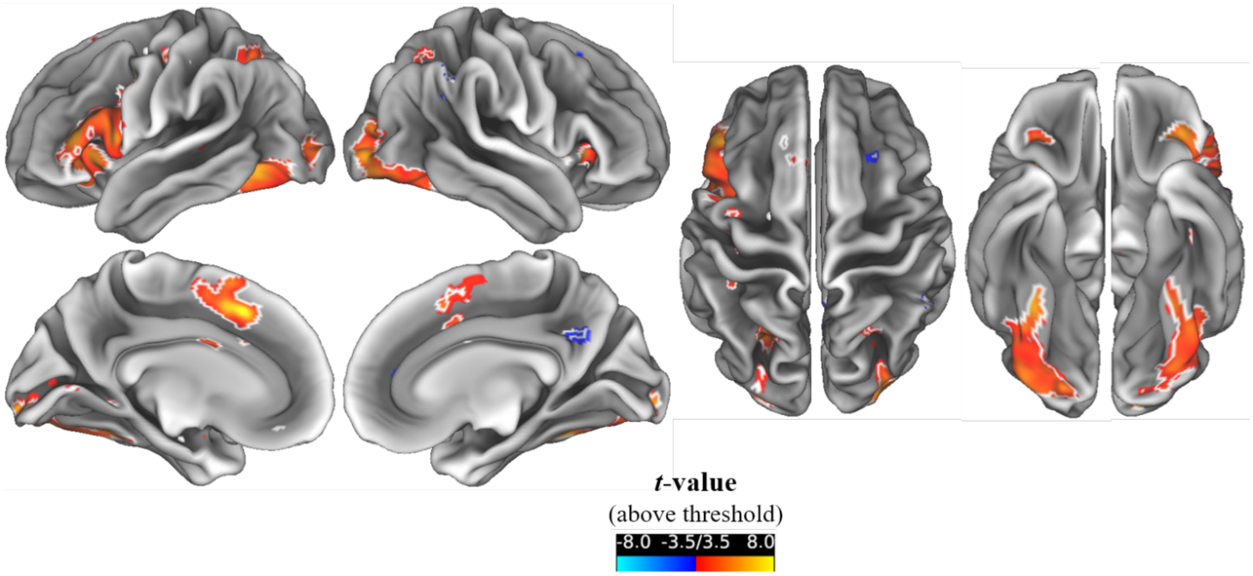

Figure 3.9. Parametric modulation of LME model residual unaccounted for by item effects (subject-driven component of encoding activation).

3.2.5 Three distinguishable components of encoding activation

Figure 3.10 illustrates all three distinctive components of SME documented above, when included in the same parametric modulation regression framework (thus, the unique contribution of each, controlling for the other two terms). When comparing to the subsequent recognition strength boundary (white outline), it is clear that the two item-driven components (red and green) implicate new regions that were not shown in the original recognition strength map. In effect, these systematic, subsequent memory phenomena were overpowered by the subject-driven component and their absence in the original subsequent recognition map reflects a potential type II error that occurs when item effects are not explicitly modelled in the design.

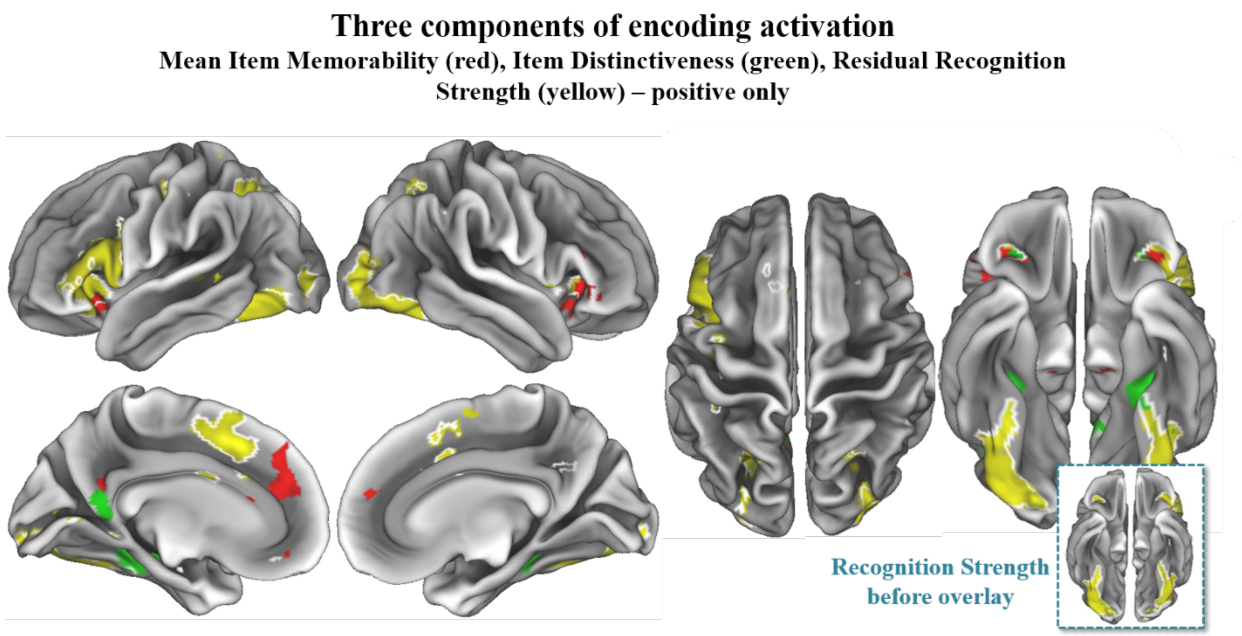

Figure 3.10. Unique contribution of each term in the LME model representing each of the three distinctive components of encoding activation.

Figure 3.11 demonstrates that several of these regions are also not present in the meta-analysis of subsequent memory for verbal items by Kim (2011). More specifically, neither the regions marked in red (left DMPFC and bilateral IFG) nor the regions marked in green (bilateral ventral MTL) are present in the meta-analysis map. Interestingly, some portion of the ventral MTL activations were implicated when the meta-analysis included subsequent memory for pictorial materials as well as associative retrieval designs (not shown). In the current study, this response was tied to item distinctiveness, suggesting that encoding of pictorial materials and associative memory might heavily depend upon item distinctiveness.

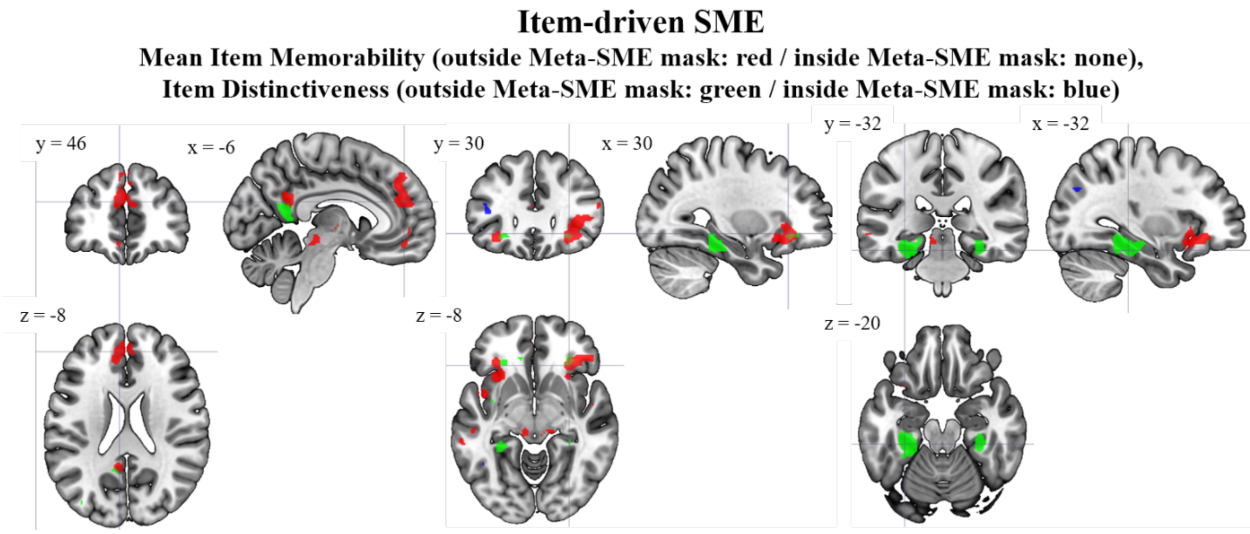

Figure 3.11. Item-driven components of encoding activation either overlapping with (blue) or independent from the meta-analysis SME map of verbal item recognition (red and green).

Table 3.8 shows the examples of the items that were predicted to be the most and the least memorable by the full LME model (again, the model is about "item" contribution to the recognition strength rating, when both "item distinctiveness" and "item memorability" components along with subject variation are jointly considered) and the items predicted by each item component separately. Note that the residual of the model is unique to each participant (because it is the difference between the model prediction and the participant's own recognition strength rating at each trial), so the residual term is unable to produce the rank ordering of items shared across participants.

Table. 3.8. Top 10 most and least memorable items predicted by the full "item effect" LME model and the two item components of the model.

Most Memorable			**Least Memorable**		
The full model prediction	**Item distinctiveness**	**Item memorability**	**The full model prediction**	**Item distinctiveness**	**Item memorability**

condom	cola	context	demise	story	demise
llama	cider	aura	leeway	concept	skirmish
mommy	locker	ether	skirmish	context	album
doughnut	goalie	enzyme	notion	effort	primate
bowel	marble	maxim	effort	extent	locker
baboon	lotion	money	anguish	series	notion
tumor	pillow	bistro	extent	logic	viper
cola	receipt	success	guidance	insight	leeway
organ	jelly	cancer	outrage	revenge	address
cancer	honey	prestige	barrage	success	anguish

3.2.6 Linking item-related neural responses to individual differences in verbal IQ

Behaviorally, I examined the moderating effect of verbal IQ on the relationship between the normative item characteristics and recognition strength responses (section 3.1.5), here I apply a similar approach addressing the question of whether the relationship between the SME activation and normative items characteristics is also moderated by verbal IQ scores. To do so, it is necessary to obtain an estimate of encoding activation on each trial for each participant within the recognition strength map. This was done using a modified beta-series analysis which fits a the effect of each trial into a separate model (Rissman, Gazzaley, & D'Esposito, 2004; Turner, Mumford, Poldrack, & Ashby, 2012). Thus, for each participant, a beta value in each trial, summarizing the activation within the entire strength map were extracted. Then, the correlation between each of the three item variables (IMG, Phono Dist and Semantic Dist) and the trial-wise betas was calculated yielding three correlations coefficients for each participant. As with the behavioral data, I then tested whether these correlations were moderated by the verbal IQ scores of the participants.

As in the behavioral analysis in Section 3.1.5, the Shipley vocabulary score reliably moderated the relationship between Semantic Distinctiveness and brain activation within the strength map across subjects (although it was unreliable when the verbal IQ outlier was excluded). In contrast, Shipley scores did not moderate the relations between activation and Imageability or Phonological Distinctiveness across the subjects, with or without the verbal IQ outlier. (Table 3.9).

Table 3.9. Correlation between individual's Shipley score and variable-to-brain correlation

Influence on activation from	With Verbal IQ outlier		Without Verbal IQ outlier	
	Pearson r	***p-value***	***Pearson r***	***p-value***
IMG	-.09	.70	-.05	.84
Phono Dist	-.15	.62	-.13	.57
Semantic Dist	**.44**	**.04**	**.30**	**.18**

The same steps were repeated for NAART35 scores. Again, the relationship between semantic distinctiveness and the brain activation was reliably moderated (but not without the outlier) by NAART35, but it was not the case for the other two item characteristics.

Table 3.10. Correlation between individual's NAART35 score and variable-to-brain correlation

Item effect variable	With Verbal IQ outlier		Without Verbal IQ outlier	
	Pearson r	***p-value***	***Pearson r***	***p-value***
IMG	.21	.33	.28	.23
Phono Dist	-.10	.66	-.07	.75
Semantic Dist	**.44**	**.04**	**.35**	**.12**

Overall, the trial-wise activations in recognition strength regions did not have significant zero order correlations with any of three normative predictors of recognition strength (not shown). However, across participants, the strength of these relationships itself was suggested to be moderated by verbal IQ scores, selectively for semantic distinctiveness. That is, subjects with higher verbal IQs showed a greater positive relationship between the trial-wise activation in the recognition strength map and semantic distinctiveness of items. Overall, these findings suggest that the degree to which semantically distinctive words produce greater activation in the SME regions (regions linked to better encoding) depends on the individual's semantic knowledge. However, this "trend" was not reliable when the Verbal IQ outlier data were excluded from the analysis, potentially due to small sample size (and lack of variability) for an individual difference analysis.

Chapter 4. Discussion

The current study aimed to isolate item-driven versus subject-driven contributions to the subsequent memory effect in functional brain imaging data. By providing a fixed set of encoding items to the participants, explicit examination of item contributions was made possible using the linear mixed effect (LME) modeling framework. In behavior, I first demonstrated that imageability, semantic and (inverse) phonological distinctiveness exerted reliable influences on each participant's recognition strength ratings at the trial-level as well as on item memorability scores at the item-level (**normative item characteristics approach**). These variables remained unique predictors of recognition strength when combined into a single multi-predictor LME model (phonological distinctiveness approached significance) which motivated their combined use in the later analyses of brain activation during encoding. In the LME model, the random intercept of item played an important role in capturing variability not accounted for by the normative item characteristic(s) considered. While demonstrating that item effects are reliable, the above analyses did not establish whether they were large or small in absolute magnitude. To address this, I explored whether consensual memorability itself can be used as a predictor for subsequent recognition performance within each individual. The leave-one-out procedure confirmed that there were reliable item memorability effects within the sample, but importantly, also demonstrated that the item effect was not the major contributor in explaining the variance within the trial-wise recognition outcomes (**item memorability approach**).

These findings validated the use of the LME model containing two item-related contributors (normative characteristics and item memorability) whose prediction for subsequent recognition strength could be used to interrogate brain activation during encoding, with the aim of separating item-driven from subject-driven (viz., subject-by-item) activations.

As will be discussed in further detail below, I suggest the first item component of the LME (the linear combination of three fixed-effect variables) isolates regions sensitive to net distinctiveness of the items and the second item component (random intercept) is linked to general item memorability. In addition to these two item components, the residual of the LME model itself was used as another predictor capturing regions sensitive to subsequent recognition strength that are unrelated to systematic item influences. As a result, the LME modeling approach produced three separable components in the SME; two of which are linked to items (hence the item-driven SMEs) and one unique to each subject (hence the subject-driven SME).

4.1 Item Distinctiveness Component

Among the seven normative item characteristics initially considered, imageability, semantic distinctiveness and (inverse) phonological distinctiveness were found to be the reliable behavioral predictors of subsequent recognition at the level of item/trials, and they remained reliable when jointly used to predict recognition strength ratings.

During fMRI, when the model prediction based on the linear combination of the three normative item characteristics was entered as a parametric regressor, it implicated regions in ventral MTL (bilateral aFuG/pPHG and left hippocampus), bilateral IFG, and bilateral posterior cingulate/precuneus. Of these, the ventral MTL region remained reliable even when other model components were also considered (Figure 3.10) Thus, in this section, I will focus my discussion on the role of ventral MTL and how it relates to the construct of item distinctiveness during encoding.

The ventral MTL, especially PHG is known to be involved in a variety of cognitive processes while being implicated during both memory tasks and non-memory tasks. In memory tasks, although the region is associated with successful encoding in general, it shows greater activation for encoding of pictorial items than verbal items (Kim, 2011) and for relational memory (e.g., associative recognition or source recollection) than item memory (e.g., item recognition or familiarity given Remember/Know judgment) (Davachi, 2006; Davachi et al., 2003; Diana, Yonelinas, & Ranganath, 2007, 2013; Eichenbaum et al., 2007; Kim, 2011). In terms of memory retrieval, the region is often associated with autobiographical memory retrieval (Addis, Moscovitch, Crawley, & McAndrews, 2004; Maguire, 2001; Svoboda, McKinnon, & Levine, 2006), which is generally assumed to be contextually rich, reflecting fairly unique experiences.

In non-memory tasks, while PHG is associated with processing of spatial information such as scene construction (Hassabis, Kumaran, & Maguire, 2007), spatial navigation (Epstein, 2008), and discrimination between the environments (Hassabis et al., 2009), it is also clearly involved in processing of non-spatial, semantic relationships that are potentially visual in nature. For example, it has been shown to be selectively sensitive for images from specific visual (semantic) categories (scenes and non-spatial objects such as faces and toys) (Diana, Yonelinas, & Ranganath, 2008) and produces greater activation during semantic tasks than non-semantic tasks specifically for words having strong visual associations (Bonner, Price, Peelle, & Grossman, 2016).

When putting the memory and non-memory findings together, it can be inferred that PHG is critical for processing relational information that is primarily but not exclusively visual, mediating both spatial and non-spatial contextual associations (Aminoff, Kveraga, & Bar, 2013;

Bar, Aminoff, & Schacter, 2008). This is consistent to the fact that two major normative predictors used to identify the region were imageability (ratings for visual imagery) and the semantic distinctiveness that was defined by relative distance between an item and the other remaining items within the vector space. This finding that items that are easily imageable and dissimilar relative to others leading to better subsequent memory also stands together with earlier behavioral findings that the combination of item-specific and relational processing leads to superior encoding (Einstein & Hunt, 1980; Hunt & Einstein, 1981). Although the current study never employed an explicit instruction to promote “relational processing” or provided any “context” for item encoding, the pleasantness judgment task during encoding might have provided an overall relational structure upon which the participants had to engage in distinctive processing for each item. Moreover, the semantic distinctiveness variable itself is inherently a variable depicting the strength of relationship among the items. The baseline relatedness among the items could have worked as an intrinsic context where certain items could be farther, thus more distinctive than others.

Overall then, the literature is consistent with the idea that this region of MTL tracks the degree to which verbal materials elicit distinctive imagery and semantic associations, which would then lead to facilitated encoding; an interpretation that fits with the co-recruitment of precuneus/posterior cingulate which is a region frequently linked to mental imagery (Fletcher et al., 1995; Ganis et al., 2004). Moreover, as I detail next, recent behavioral research converges on the hypothesis that items vary in distinctiveness conceptualized in this manner.

More specifically, a recent study by Cox and colleagues (Cox, Hemmer, Aue, & Criss, 2018) reported a similar behavioral finding. Similar to the current study, the main purpose of this large-scale exploratory study was to examine how performance on different memory tasks is correlated with respect to the processes engaged by participants and the information conveyed by the items. The inter-task correlational structure among the five different tasks (four episodic memory tasks and one lexical-decision task) was assumed to help disentangle the role played by the information and the processes acting on that information in human memory. Specifically, the inter-task correlation between two tasks across participants (collapsing across items) would indicate the "processes" the tasks share with each other, whereas the inter-task correlation across items (collapsing across participants) would indicate the "information" that supports the performance on both tasks. Based on a reliable correlation among episodic memory tasks (single item recognition, associative recognition, cued recall and free recall), they argued that the different memory tasks involve similar memory structure (process and information) that is simply accessed in different ways, which, however, is distinct from lexical decision.

Critically, the results from their principal components analysis (PCA) were particularly relevant to the current study. The PCA revealed latent dimensions reflecting how item information contributed to performance. These item-related dimensions showed reliable correlations with two normative word characteristics for episodic memory for words; concreteness rating norms (Brysbaert, Warriner, & Kuperman, 2013) and semantic specificity, which they defined as the average dissimilarity between the documents in which a word appears. In discussion, they argued that the two factors are related to "distinctive semantic features" elaborating as below.

> Here, "distinctive" semantic features means that a word refers to a specific concrete entity, and is thereby associated with perceptual features of that entity (Paivio, 1969), and/or it is used only in specific discourse contexts (low semantic diversity) and is therefore associated with a narrow set of patterns of use (Adelman et al., 2006).
>
> (p. 567).

Although "concreteness" is not equivalent to "imageability", the two variables are often used interchangeably (Binder, Westbury, McKiernan, Possing, & Medler, 2005; Sabsevitz, Medler, Seidenberg, & Binder, 2005) and they were found to be highly correlated for nouns (Paivio, Yuille, & Madigan, 1968). In fact, the imageability norms and the concreteness ratings taken from Brysbaert, Warriner and Kuperman (2013) demonstrated a robust positive correlation ($r = .78$, $p < .0001$) for the current encoding material. This suggests that in the current study, it may have been easier to form a distinctive mental image for the concrete nouns than abstract nouns. This in turn would also explain the precuneus activation given the well-replicated link between precuneus activation and mental imagery or concreteness in the imaging literature (Cavanna & Trimble, 2006; Fletcher et al., 1995; Fliessbach, Weis, Klaver, Elger, & Weber, 2006; Ganis et al., 2004).

Taken together, it can be concluded that the combination between concreteness/ imageability and semantic distinctiveness (along with low phonological distinctiveness reflecting easier/more common pronunciation for fluent processing) is a good recipe for distinctive processing of words which in turn leads to durable encoding. Moreover, highly imageable and semantically distinctive items may have benefited from dual-coding (Paivio, 1971; 1986) where the vividness of perceptual representation and distinctiveness of verbal representation, in this case, would jointly benefit encoding. This also provides a potential explanation for the similarity between the current SME map and the meta-analysis SME map for pictorial material (Kim, 2011), both of which demonstrated extensive occipito-temporal activations.

In the current study, I have interpreted the joint prediction from the three normative item characteristics as capturing item distinctiveness. This formulation of distinctiveness is more abstract than possible by focusing solely on any single item attribute. Under this formulation for example, an item might end up with a moderately high net distinctiveness value because it is highly imageable, but only modestly semantically distinctive. However, another item might have this same level of item distinctiveness because it is highly semantically distinctive, yet only modestly imageable. Of course, an item high in both attributes would have even higher net distinctiveness. Thus, the item distinctiveness construct I applied to the three-variable term is abstract in the sense that it reflects the net distinctiveness of a set of functionally independent attributes. This is broadly consistent with the notion of distinctiveness in episodic memory since the distinctiveness of an experience is unlikely to reflect merely the uniqueness of any single attribute. Instead, distinctive experiences typically reflect a unique collection of event attributes relative to one's other experiences.

To sum up, the weighted combination of three item characteristics from the fixed effect portion of the LME model tracking recognition strength behavior, also uniquely predicted encoding activations in bilateral ventral MTL that were **not** part of original recognition strength map. The ventral MTL activations were anterior to those identified in the original strength map and I have interpreted the response as reflecting the net distinctiveness of the verbal items. Consistent with this idea, PHG is known to process co-occurring multisensory inputs (Diaconescu, Alain, & McIntosh, 2011) as well as co-occurring items converging within a context (Aminoff et al., 2013). This raises the possibility that as one moves more anteriorly along the ventral MTL surface, activations will track the multidimensional distinctiveness of encoding experiences in increasingly abstract ways.

4.2 Item Memorability Component.

The second item-driven component was linked to the average of the recognition strength ratings that each item produced across the participants. This effect was estimated by the random item intercept term of the LME model capturing the tendency of each item to elicit a particular level of strength rating controlled for inter-subject differences in the rating and the item distinctiveness effect discussed above. The random item intercept component closely tracks simple item memorability estimates calculated from the sample (i.e., correlation with the average strength rating of each item across the subjects, $r = .95$; correlation with consensual hit rates from the sample, $r = .90$), and conceptually both reflect item effects, over and above the normative characteristics specifically considered in the current report.

Importantly, when item memorability was used to predict the trial-wise outcomes in each subject's behavior, the majority of participants showed a small but reliable relationship between item memorability and their recognition decisions (section 3.1.4). This demonstrates that although general item effects are present in behavior, the vast majority of each subject's recognition behavior is **not** predictable on the basis of how others respond to those same items. In addition, there was only a modest correlation between the item memorability estimates in the current study and those measured in Cortese et al. (2015), suggesting that the effect of verbal item memorability may not generalize beyond the context/research design within which it was measured.

In brain imaging based on the LME model predictions, the random item intercept term implicated a set of regions that partially overlapped with the item distinctiveness component discussed above. While bilateral IFG and left posterior cingulate/precuneus activations were shared between the two item components, a robust left DMPFC activation was unique to the item intercept component. The left DMPFC activation is also well outside the boundary of original recognition strength map (Figure 3.10, area in red) resulting in the isolation of a second item-driven SME largely focused within the PFC (in addition to the bilateral ventral MTL, the item distinctiveness effects discussed in the previous section).

The functional contribution of this region to item memorability effects is difficult to ascertain because of its proximity to anterior cingulate cortex (ACC), which is implicated in many cognitive tasks thought to require conflict monitoring (Botvinick, Cohen, & Carter, 2004; Kerns et al., 2004). As an exploratory test of the "conflict monitoring" hypothesis, I calculated the degree of consensus of pleasantness ratings to each item under the assumption that low consensus items would be associated with high decision conflict. However, the consensus values were unrelated to the item intercept estimates ($r = .06$, $p = .25$).

A second possibility is that this activation is linked to the so called "core network" which is active during tasks involving self-referential thinking, theory of mind, self-projection, and autobiographical memory retrieval (Buckner & Carroll, 2007; Gusnard, Akbudak, Shulman, & Raichle, 2002; Isoda & Noritake, 2013). This network of "me-ness" is known to be closely overlapping with (or equivalent to) the default network (Buckner et al., 2008; Raichle et al., 2001). Under this interpretation, the DMPFC activation might reflect the subjects' tendencies to use autobiographical episodes or self-relevant semantic knowledge when rendering pleasantness judgments. However, it must be kept in mind that the item intercept estimates, by construction, is an item-based value that spans subjects. This interpretation would require stable differences across the items in the degree to which they elicit retrieval of self-relevant autobiographical or semantic information during evaluation. As an initial test of the face validity of this idea, one can consider Table 3.8 which lists the 10 most and least memorable items under the different LME model component scores. The most memorable items under the item intercept component do not appear particularly self-referential providing little support for this interpretation. Thus, while the DMPFC activation is clearly outside the basic SME map, and hence was only identified by specifically modeling item influences, its functional role is unclear.

4.3 LME Model Residual: Subject-driven SMEs

A major implication of the current dissertation is that the subsequent memory effect reported in the literature is not attributable to item effects. After controlling for possible item-related effects that were **measurable within the current study**, the non-item-related residuals of the LME model recovered the original recognition strength activations. This specifically indicates that the original SME strength map (Figure 3.1) is not a function of item effects that span participants. Moreover, the parametric analysis using the LME residuals not only recovered the original SME map, it slightly expanded its borders.

Along with the fact the item components, defined by the model, were clearly outside the original map (i.e., bilateral ventral MTL and left DMPFC in particular), the expansion of the LME residuals map compared to original indicates that the LME modeling approach was a more powerful way to identify subsequent memory effects in general. To see why, it is necessary to assume there may be two types of SME effects in general. First, there are presumably responses to items that are subject-specific. For example, when presented with the word ‘puppy’, it is likely the case that a subject who just adopted a puppy would find the encoding experience particularly memorable given the extra-list associations the probe evoked. Such a subject-specific processes cannot be modelled simply by knowing the particular items a subject is viewing and they constitute what is formally known as subject-by-item interactions. In contrast, if there are items that tend to drive similar processes across participants, this can be only demonstrated by modeling/estimating an item effect that spans participants. In the case of 'puppy', subject-driven effects aside, this word might be moderately memorable on average in the population. However, when using only the raw behavioral responses of the participants, these two types of effects are conflated within a single response to each item. This necessarily reduces the efficiency to detect either effect. The relative cost of this conflation depends upon the strength of the two effects, and since the item-driven effects were small (though reliable), they were the ones not detected when using the raw responses as fMRI covariates. Since the LME modeling approach removes (estimated) item components from the subject's responses, it improves the ability of what is left of those responses (the residuals of the item LME model) to detect subject-driven phenomena **and** enables the detection of item-driven phenomena. That is, it increases power.

4.4 Design Differences and Reliability of Consensus as a function of Sample Size

As noted previously, the Cortese study (2015) and the current study conducted on a subset of the same verbal materials demonstrated two potential discrepancies. First, the studies differed in which normative word characteristics explained the item memorability measures and the total amount of variation in these measures accounted for. Second, the item memorability scores themselves were only modestly correlated across the two studies (r = .38).

One possibility is that both discrepancies are the result of measurement reliability issues. In particular, the current item memorability estimates are based on a smaller sample (N = 22) than those of Cortese et al. (N = 60) and this may limit the degree to which we should expect close correspondence. However, there are several arguments against this interpretation. First as shown in section 3.1.4, the item memorability scores in the two studies were similarly useful in predicting the individual responding of subjects (mean r = .11 versus mean r = .16). Second, the item components derived from the LME in the current study were sufficiently reliable to yield reliable brain activations in regions not detected by the original SME model (Figure 3.10). Since brain activations themselves are noisy phenomena, this would be unlikely if the estimates were also noisy. Finally, when directly compared as imaging covariates during encoding, the sample item memorability scores and those of Cortese et al. yielded similar activation maps (not shown). Had the Cortese et al. estimates been more reliable indicators of a common effect present in both studies, they would have yielded a more robust activation map. These outcomes suggest that the difference in the item memorability-based outcomes may not be primarily due to estimate reliability but may instead reflect the fact that verbal item memorability is sensitive to study design.

In current dissertation, with the aim of explicitly examining the item effects within the subsequent memory paradigm, several choices were made to maximize potential item effects in a way that would produce a greater commonality across participants. For example, whereas the Cortese et al. (2015) study did not adopt any orienting task during encoding, the current study used pleasantness judgments for words combined with lexical decisions by mixing the words with nonwords trials. The unstructured, intentional encoding of the former may emphasize the role of item-driven phenomena if subjects vary considerably in their approach and attentional engagement with the task, whereas providing an orienting task may enforce more common processing demands that lessen the role of item-driven phenomena.

Since the current study was also interested in semantic distinctiveness as a contributor to encoding, and semantic processing is a dominant contributor to episodic memory, I adopted the pleasantness rating task for orienting. This task is also preferred for large item pools (with a single encoding-test cycle for 400 items), since it will yield good subsequent recognition performance even in the face of considerable proactive interference during testing. Regardless, as mentioned above, when using the leave-one-out procedure, section 3.1.4 demonstrated that item memorability estimates accounted for a similar proportion of trial-wise behavior in the current data and the Cortese et al. Thus, the use of the pleasantness orienting task did not appear to eliminate item effects per se.

Although only two studies are considered, this raises a general possibility that the nature of encoding doesn't eliminate item effects per se, but instead changes the features that dominate item effects. For example, deep-encoding tasks may attenuate the effect of non-semantic variables (e.g., word frequency), leaving the contribution of more semantic variables relatively intact. If this hypothesis is correct, the manipulation known to produce better encoding (potentially by encouraging deeper processing) such as survival processing or animacy judgment (Nairne, Thompson, & Pandeirada, 2007; Nairne, VanArsdall, & Cogdill, 2017; Nairne, VanArsdall, Pandeirada, Cogdill, & LeBreton, 2013) may further reduce the item characteristics contribution of lexical variables. Also, the nature of encoding (and retrieval) may reliably alter the rank ordering of items in terms of aggregate item memorability, which would be quite compatible with seminal demonstrations of encoding specificity (Thomson & Tulving, 1970; Tulving & Thomson, 1973) and transfer appropriate processing (Kolers & Roediger, 1984; Morris et al., 1977) in the literature. Interestingly, despite the simplicity of this question, this dissertation appears to be the first study to actually compare item memorability estimates, for matched items, across different research designs.

One final design choice in the current study that may have weakened the item memorability relationships across the current design and that of Cortese et al. (2015) was the intermixing of the lexical decision task and the semantic decision judgments here. As noted in section 2.2.2, this was done to provide a non-semantic stimulus class that might help the interpretation of SMEs later discovered. However, it also means that to some extent, explicit lexical decision-making might have rendered certain item characteristics more salient or potent during encoding compared to the Cortese et al. procedures. For example, orthographic distinctiveness was not predictive of item memorability in the current study whereas it was highly reliable in Cortese et al. (2015). If so, however, this would emphasize the broader point that item memorability is not a fundamental attribute of items but instead the result of how items may interact with encoding goals and processes.

4.5 Is Memorability an Intrinsic Normative Property of Items?

As noted above, the comparison of item memorability estimates across the current study and that of Cortese et al. (2015) suggests that item memorability is malleable and depends upon encoding procedures (and presumably their interaction with retrieval procedures). However, in a number of studies investigating memory for several categories of visual stimuli (e.g., faces, scenes, etc.), Bainbridge and colleagues argued that memorability is an intrinsic property of items that serves as a bridge between perception and memory (Bainbridge, 2016; Bainbridge et al., 2017; Bainbridge, Isola, & Oliva, 2013; Bainbridge & Rissman, 2018; Bylinskii, Isola, Bainbridge, Torralba, & Oliva, 2015).

Similar to the current study, a major goal of Bainbridge and colleagues (Bainbridge et al., 2017) was to isolate the neural signature of item memorability that may have been confounded with SMEs estimated in prior studies. In their approach, item memorability was estimated from a large-scale online recognition memory study with each item score reflecting the proportion of online sample participants garnering hits to each item. They then compared the univariate memorability contrast comparing high versus low memorability items during encoding, versus the traditional SME contrast comparing subsequent hits versus misses to those same items in the study. Critically, they identified areas within the ventral visual stream and MTL as specific to the memorability contrast, concluding that the MTL signals a high-level perceptual property of stimuli linked to their canonical memorability. Since the majority of these regions did not show reliable subsequent memory effects (greater activation for subsequent hits versus misses), the authors argue that the regions do not regulate the encoding of the item into memory per se, which was instead held to critically depend upon left lateral PFC; a region demonstrating reliable subsequent memory contrast effects. The fact that memorability and subsequent memory demonstrated separable neural correlates led the researchers to conclude that the two memory-related constructs were dissociable from each other. The researchers argued that subsequent memory effect that had been often found in both the MTL and PFC may be made up of two separable components, memorability in the MTL and individual subsequent memory in PFC.

The stimuli used in Bainbridge et al. (2017) and the current study fundamentally differ (words versus face and scene images) and so it is perhaps not surprising that the studies reach different conclusions. However, the divergence in the findings and conclusions is noteworthy. For example, the current study identified both item-driven and conventional subsequent memory effects in PFC, but in different areas. As Figure 3.10 shows, there is a robust left ventrolateral

PFC response for subsequent memory even when controlling for potential item-driven effects (LME residuals covariate). In contrast, Figure 3.7 demonstrates there are several other regions (e.g., DMPFC) of PFC sensitive to the item intercept term in the model, which is essentially a proxy for item memorability estimated within the sample. Thus, the current data do not point to a sole role for PFC in either subject- or item-driven subsequent memory phenomena.

Turning to areas of the ventral visual stream and MTL, the data demonstrate the analogous finding that the region supports both item- and subject-driven subsequent memory phenomena. Whereas the LME residuals component isolates subject-driven SMEs to bilateral fusiform extending forward into PHG, the item distinctiveness component of the model identified bilateral MTL responses anterior to this (areas in yellow versus green in Figure 3.10). Since the latter was coded as a fixed item effect, whereas the former was coded as the portion of recognition strength responses that cannot be explained via item effects in general, the data demonstrate functionally different responses.

As noted above, one key difference across these studies is the use of words versus pictures. With respect to the words, it seems highly unlikely that strictly perceptual phenomena are the dominant contributor to recognition encoding even though prior work demonstrates that orthographic distinctiveness can play a role (Cortese et al., 2015; Glanc & Greene, 2007; Kirchhoff, Schapiro, & Buckner, 2005). Nonetheless, relative to images of faces and scenes, it is clear that words are perceptually impoverished. Given this and the use of a deep processing task in the current study, much of the information driving recognition outcomes is likely semantic in nature, having little to no link to the features of the materials below the lexical level. Indeed, the dominant normative characteristics for predicting item memorability in the current study were non-perceptual in the sense that they are not a function of the sub-lexical features of the stimulus.

From this perspective, one can expect that the neural correlates of memorability should be differentially defined per each stimulus type.

Despite the differences in the two studies, they converge in that Bainbridge and colleagues also suggested that MTL regions were critical in coding for ‘multidimensional distinctiveness’. However, in that report the memorability is hypothesized to be “a high-level perceptual property reflecting the statistical distinctiveness of a stimulus along a multidimensional set of axes, beyond a simple single measure like physical distinctiveness of points in a face (p. 149)”. In this conceptualization, memorability, the probability that others will recognize the item in a large-scale normative study is considered a proxy for the statistical distinctiveness. In contrast, item distinctiveness in the current study was based on stimulus properties that are independently estimable outside of memory findings. For example, the semantic distinctiveness scores of the items do not depend on normative studies of recognition. Instead, they are defined by the distribution of the words across a large corpus of texts (see Appendix I). Conceptually, items are held to be semantically distinctive to the degree they tend to have a unique distribution across these texts and hence are unlikely to share a meaning with the other items. This is important, because the manner of definition does not guarantee that the scores will in fact systematically predict recognition memory outcomes. Analogously, the imageability scores are derived in a fashion that is independent of recognition memory testing, and so these too could fail to predict outcomes. Finally, as with the semantic distinctiveness, there is a psychological reason as to why one might anticipate that imageable items form more distinctive encoding experiences. In contrast, when ‘distinctiveness’ is operationalized as the probability that individuals will recognize an item when studied, circularity becomes a concern.

4.6 Item Effects are Small during Verbal Recognition

Despite the fact that even small item-driven effects can be theoretically important, it is important not to overstate their contribution to recognition behavior. As noted in section 3.1.4, the efficacy of the consensus of others in predicting an individual's responses during recognition was quite limited, accounting for approximately 1% to 2% of performance. Nonetheless, the effects are quite statistically reliable because the same small contribution is present in most observers. While memory researchers might find this small contribution of item effects surprising, this may reflect the various ways the term 'item effect' is used or conceptualized.

Again, as I noted in section 3.1.4, in its simplest form, an item effect reflects the ability to predict subjects' recognition responding by knowing the items he or she is being tested on. To say an item effect is large in this sense, is to claim much of the subjects' behavior can be predicted given knowledge of the items. However, as noted above, from this perspective, the vast majority of the subjects' behavior cannot be explained by knowing the items he or she is being tested on **when using consensus** to operationalize memorability. Although consensus measures may seem a common sense approach, they can only work when item-linked representations and experiences are highly shared across individuals. In the current study, the preliminary finding that verbal IQ may moderate the link between the ability of semantic distinctiveness to facilitate encoding (section 3.1.5) begins to suggest there are boundary conditions to the utility of consensus measures. Put simply, individuals with increasingly less semantic knowledge would be expected to show increasingly fewer mnemonic benefits from semantically distinct materials. Analogously, while words linked to expertise in particular domains might be semantically distinctive to those experts (e.g., chemists, statisticians, etc.), they would be less so for novices (Long & Prat, 2002). Critically, this does not mean that the mechanisms or processes underlying distinctive encoding are absent for novices, or individuals with low semantic knowledge, but that the scores reflecting the degree of distinctiveness of each item are less calibrated to the structure of these individuals' semantic knowledge.

Aside from boundary conditions linked to individual differences, the low predictive power of item memorability should not be confused with questions of reliability or the degree to which one can explain variation in item memory scores. For example, in the case of memorability for images in Bainbridge et al. (2017), the internal consistency of the memorability scores can be quite high reaching a mean of ρ = .69 for faces and ρ = .75 for scenes across random split halves for a sample of 800 participants. In other words, when one divides the sample into two groups and calculates the proportion of subjects that successfully recognized each item, the scores of the two sub-samples are highly correlated. However, this does not mean that these scores will necessarily predict the trial-wise behavior of individuals well, even if they are used in testing designs that resemble the manner in which the scores were derived. To see why, consider item memorability scores in the range of .6 to .4. While these scores may have high split-half reliability in a large sample, the absolute values of the scores indicates that there is considerable uncertainty around whether the items are likely to be recognized or forgotten for any given individual (roughly, given each item, a half of the observers would say 'old' and the other half would say 'new'). Nonetheless, because there are numerical differences in memorability across items within this small range, the rank order across the items can be still formed and this order can be quite consistent across the split-haves if the sub-samples are large enough to reliably resemble the original. Thus, with a large sample size that can yield reliable memorability estimates, the split-half correlation will be always quite high, because the measure indicates the strength of the association between the ranks of memorability scores, regardless of the variability within the scores.

Thus, a high split-half reliability does not assure that the items convey information (viz., variability) relevant for predicting behavior. Rather, it is a statement that the particular measurement has enough observation to demonstrate a reliable rank order across items. On the contrary, the consensus prediction adopted in the current study can be more relevant construct for explaining behavior rather than representing the measurement property.

4.7 Limitations and Future Directions

The current data suggest that much of the activation typically ascribed to subsequent memory effects is subject-driven. That is, it is not predicted by knowing the items the subject is considering, at least as captured by the LME modeling framework that I used. However, this conclusion does not mean that there might not be better ways of isolating item effects. The fact that the item components of the LME model implicated additional areas outside of the original SME map demonstrates these item components are useful and sufficiently reliable to detection encoding activations (which are quite noisy themselves), however, the findings may be fairly dependent upon the design, which in this case used deep processing with a brief retention interval. This raises questions with respect to whether item-driven contributions should decline or increase, and likewise whether subject-driven contributions should decline or increase with other manipulations.

In the case of the current design, only one of the original variables considered by Cortese et al. (2015), imageability, was predictive of consensual item memorability scores. Additionally, semantic distinctiveness scores which I derived specifically for this study, were also predictive of item memorability scores in the two data sets but those two were the only overlapping variables reliable in both studies. Moreover, the item memorability scores from the two data sets were only modestly correlated. While I have interpreted this modest correspondence across the data sets as reflecting differences in the designs, primarily linked to encoding differences, this hypothesis remains to be directly tested.

In terms of the variables chosen, even though this dissertation examined a comprehensive set of lexical and semantic variables introduced by recent mega-study (Cortese et al., 2015), there are several other variables that were not examined but are potentially important. For example, semantic distinctiveness as defined in the current study was based on a simple type of vector semantics computation conducted on a specific corpus. With recent advancements in natural language processing and multivariate analysis techniques in psycholinguistics, various ways to quantify semantic distinctiveness have been introduced and critically, each approach is based on its own unique operational definition of what the "distinctiveness" of a word is. For example, contextual diversity (Adelman, Brown, & Quesada, 2006) is based on number of contexts (passages or documents which a word appears) calculated from a corpus whereas semantic diversity (Johns, Sheppard, Jones, & Taler, 2016; Jones, Johns, & Recchia, 2012) takes the information redundancy of the contexts into consideration and uses document similarity as a modulating factor for number of contexts count. Whether semantic distinctiveness quantified by a different approach would provide similar activation patterns as the current variable is another interesting research question.

Moreover, there are other, basic semantic variables that might contribute to subsequent memory. Instead of "distinctiveness" which represents the distance within a context between the words, we can also quantify semantic aspects of the words based on how people approach the words given a task. These task-based indices can be normed from a variety of semantic tasks such as semantic association generation task or feature-listing task. For example, Pexman and colleagues (Pexman, Hargreaves, Edwards, Henry, & Goodyear, 2007) defined the semantic richness of words using the number of associates people generate to a word. When asked to report the first word comes to mind given a cue word, a group of people can provide a common answer (low number of associates; low semantic richness) or each person can provide an unique answer so that many different answers could be generated from a group (high number of associate; high semantic richness). Using this measurement, Pexman and colleagues found that semantically richer words were judged faster in behavioral tasks such as word naming and semantic categorization. During functional imaging, semantically richer words produced lower activation than words with lower number of associates in regions such as cuneus, left IFG and left ITG. They concluded that faster semantic settling for words with rich semantic representation lead to lower activation in these regions. An important future question would be to see whether the semantic richness score measured in this way correlates with semantic distinctiveness measure and to consider relative contribution of the two to subsequent recognition strength. Because the semantic distinctiveness scores emphasize uniqueness, but the semantic richness scores emphasize diversity of associative features, they may not implicate similar regions.

Another fruitful direction might be extending the fMRI analyses to multivariate approaches. The current proposal used strictly univariate analyses, which is appropriate given that the SME effects in the extant literature are a univariate phenomenon. However, the method I used to quantify semantic distinctiveness of words bears a clear similarity to representational similarity analysis (RSA) in functional imaging. In the case of semantic distinctiveness, the pairwise dissimilarity of occurrences across a corpus of texts defines the distinctiveness of the words. Analogously, one can examine the pairwise similarity/dissimilarity in the pattern of activations across voxels within a selected region to see if it adds additional predictive value in terms of subsequent memory outcomes. For example, such a pattern analysis performed on the regions in the bilateral MTL demonstrating the current item distinctiveness component could be tested for incremental validity (but see LaRocque et al. (2013) for potential discrepancy in patterns across sub-regions within the MTL).

Finally, for the linguistic variables considered, only linear/monotonic effects of the variables were examined in the current study. Researches have shown a quadratic effect of length (Yarkoni et al., 2008; Schuster et al., 2016) and word frequency (Hemmer & Criss, 2013) on recognition memory. Thus, there might be a more effective way of modeling these item variables in predicting recognition strength as well as in targeting regions sensitive to these variables during functional imaging.

4.8 Conclusions and Implications

Overall, this dissertation demonstrated that there are reliable but small (in absolute terms) item effects present in recognition strength data. These effects, in part, were linked to three normative variables that are consistent with the working hypothesis that net item distinctiveness facilitates recognition. When modeled via LME regression with an item random intercept term, the model-based estimates distinguished item-driven from subject-driven SMEs introducing a methodological framework that may be useful for isolating brain activation linked to different construct in other domains. Critically, the item-based components detected new regions not discovered by the conventional approach, and in particular, bilateral ventral MTL that may indicate the facilitated encoding of materials that are distinctive in a multi-attribute fashion. Moreover, the modeling of these item-driven phenomena enabled the use of the model residuals, which reflect subject strength ratings that are not explained by item effects. These residuals yielded a more robust SME map than the standard approach, presumably because they did not conflate subject- and item-driven mechanisms.

Appendices

Appendix I. Vector semantics approach to calculate semantic distinctiveness

To develop a quantitative measure of semantic distinctiveness as an item characteristic, I turned the tools of vector semantics that is similar to Latent Semantic Analysis (Landauer & Dumais, 1997). Based on the distributional semantics framework (for a review of different models of distributional semantics, see Mandera Pawełand Keuleers & Brysbaert, 2017; Rohde, Gonnerman, & Plaut, 2006; Turney & Pantel, 2010), I assume the semantic similarity of words is reflected in co-occurrence patterns discernable in large samples of natural language data. More specifically, words distributed similarly across a collection of language samples are assumed to convey similar meanings.

I used a large database of informal movie reviews culled from the Internet Movie Database (IMDb) gathered by Maas et al., 2011. This data set contains 100,000 user submitted reviews, half of which were labeled as positive or negative. The original purpose of the data set was sentiment analysis, but here I use it as a large corpus for calculating word similarity scores. An example of one of the reviews is below:

> Blistering black comedy co-written by Jill Sprecher (who also directed) and Karen Sprecher, "Clockwatchers" gives us a suffocating office setting so vivid and real I half-expected my own co-workers to show up in it. Toni Collette plays the new temporary in a nondescript building wherein office-incidentals are slowly disappearing from the supply cabinet. The ensemble acting is delightfully accurate, and the strife which ensues in this scenario is comically overwrought and horrifying. Sprecher's direction is focused and brave (no overtures to broadly comical sensibilities), and she nimbly stretches the film's satirical edge quite far without faltering. The movie is a genuine American original, and by the final third I couldn't wait to see it again from the start. ***1/2 from ****

The text2vec analysis package (Selivanov, 2016) and quanteda (Benoit et al., 2018) in R were used for the analysis. The reviews were first transformed into a term-document matrix (tdm) in which each row constituted a unique word within the collection and each column a movie review. During construction of the tdm, each review was stripped of whitespace, punctuation, numbers, and all words were converted to lower case. Additionally, common English stop words such as conjunctions, articles and prepositions were also removed. Each cell in the matrix indicates the frequency of each word's occurrence within each document (i.e., movie review). I then reduced the size of the tdm by removing all words that were not present in the 3,000 word set of Cortese et al. (2015) yielding a tdm consisting of 2,954 words (rows) and 100,000 columns.

Also, the word 'movie' was removed from the analysis even though it was in the Cortese set because it was over-represented in the database given the corpus consisted of movie reviews. In order to calculate a word's semantic similarity, I relied on the cosine similarity metric. After transposing the tdm to a document-term matrix (dtm), each column represented a target word, and the cells below each word reflected the word's distribution across the 100,000 documents in units of frequency. These vectors were then normalized using the l2 norm which transforms the vectors' cell frequencies by dividing each by the square root of the sum of the squared frequencies of all the cells in that vector (a Euclidean magnitude). Similar outcomes were obtained with non-normed raw frequencies. Thus, each vector represents the position of a word in a normalized N-dimensional space of documents, and the cosine similarity metric for any pair of vectors is their dot product divided by the cross product of their lengths.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

Thus, the semantic similarity of any pair of words is captured by the angle between their vector representations in the space. A perfectly similar pair of words would be distributed identically across the documents and have a cosine similarity of 1. A perfectly dissimilar pair of words would share no co-occurrence frequencies across the documents and have a cosine similarity of 0. Since there are no negative values possible in frequency data, the cosine similarity measures are restricted between 0 and 1. To facilitate interpretation given our interest in semantic distinctiveness, I transformed the cosine similarity into a dissimilarity score by subtracting it from 1.

Within this vector space representation, a distinctive word is one that is highly dissimilar to all the remaining words (i.e., has uniformly low cosine similarity values when compared to all other words). To estimate a word's semantic distinctiveness, I then calculated the mean of its pairwise dissimilarity values with all the remaining words in the recognition test set. Thus, words with high scores are more semantically distinctive with respect to the remaining memoranda than words with lower scores.

Appendix II. SPM result tables for suprathreshold activation maps

(uncorrected *p* < .001, 5 voxels, coordinates in MNI space)

Appendix II.1. Regions showing significant parametric modulation of Recognition Memory Strength in Figure 3.1.

Approx. AAL	Approx. BA	Voxels (kE)	peak T	peak equiv Z	x (mm)	y (mm)	z (mm)
L Supp_Motor_Area	8	1142	8.93	5.68	-6	20	46
L Supp_Motor_Area	6		6.14	4.59	-4	16	60
L Supp_Motor_Area	6		5.39	4.22	-2	8	56
L Frontal_Inf_Tri	45	3252	8.12	5.41	-46	30	10
L Frontal_Inf_Orb	47		7.89	5.32	-36	34	-14
L Frontal_Inf_Oper	44		7.29	5.09	-50	12	10
L Occipital_Mid	VisualAssoc (18)	5991	7.67	5.24	-20	-98	-2
R Fusiform	Fusiform (37)		7.37	5.12	36	-44	-20
R Calcarine	VisualAssoc (18)		7.16	5.04	22	-96	-2
R Insula	45	392	7.39	5.13	36	28	0
R Frontal_Inf_Orb	47		6.6	4.8	36	36	-8
R Frontal_Inf_Tri	46		3.99	3.4	40	34	6
R Cingulum_Mid	8	56	6.1	4.58	16	12	38
L Occipital_Mid	39	524	5.69	4.38	-26	-64	40
L Parietal_Sup	7		5.38	4.22	-20	-60	50
L Parietal_Sup	7		4.82	3.91	-24	-50	44
R Cingulum_Mid	24	43	5.45	4.26	10	4	30
L Cerebelum_4_5	Thalamus (50)	535	5.43	4.24	-4	-22	-12
L Hippocampus	Putamen (49)		5.11	4.07	-28	-24	-2
L Thalamus	Thalamus (50)		4.95	3.98	-14	-20	2
R Thalamus	Thalamus (50)	258	5.37	4.21	22	-12	2
R Thalamus	Thalamus (50)		5.02	4.03	18	-18	14
R Thalamus	Thalamus (50)		4.76	3.88	8	-14	-8
L Cingulum_Ant	24	54	5.29	4.17	-4	2	28
L Fusiform	Parahip (36)	111	5.25	4.15	-38	-12	-26
L Amygdala	Amygdala (53)		5.05	4.04	-34	0	-22
L Frontal_Sup	8	44	5.24	4.14	-16	34	48
R Angular	7	355	5.17	4.1	24	-46	42
R Parietal_Inf	7		4.31	3.6	26	-50	50
R Angular	39		4.2	3.54	30	-62	46
R Pallidum	GlobPal (51)	49	4.91	3.97	14	4	0
R Insula	44	96	4.89	3.95	48	6	4
R Insula	Insula (13)		4.29	3.59	40	14	-2
R Frontal_Inf_Oper	44		3.95	3.38	50	14	-2
L Lingual	VisualAssoc (18)	39	4.75	3.87	-18	-44	-2

L Cerebelum_3	Fusiform (37)	11	4.54	3.75	-6	-34	-24
L Precentral	PrimMotor (4)	106	4.49	3.72	-36	-16	50
L Postcentral	PrimSensory (1)		3.98	3.4	-38	-24	42
L Parietal_Inf	40		3.7	3.21	-48	-28	48
L Insula	Insula (13)	19	4.43	3.68	-34	8	0
L Cerebelum_Crus1	Fusiform (37)	12	4.4	3.66	-44	-54	-30
R Precentral	6	12	4.34	3.63	58	-2	42
L Parietal_Sup	7	7	4.32	3.61	-16	-44	74
L Parietal_Inf	PrimSensory (1)	42	4.27	3.58	-40	-36	42
L Rectus	11	24	4.26	3.57	-4	40	-16
L Lingual	VisualAssoc (18)	26	4.21	3.54	-2	-66	6
R Frontal_Inf_Tri	45	12	4.12	3.49	40	24	14
R Frontal_Inf_Tri	46	10	4.11	3.48	52	36	14
R Hippocampus	Parahip (36)	5	3.97	3.39	22	-20	-14
R Occipital_Mid	19	18	3.96	3.38	28	-68	24
L Rolandic_Oper	PrimMotor (4)	13	3.93	3.36	-38	-6	16
R Fusiform	Parahip (36)	13	3.91	3.35	36	-6	-28
R Frontal_Inf_Oper	44	24	3.79	3.27	40	8	30
L Precentral	6	7	3.73	3.23	-28	-10	60
R Hippocampus	Hippocampus (54)	6	3.71	3.22	32	-8	-16
R Cingulum_Ant	32	6	3.69	3.2	14	20	28
R Putamen	Putamen (49)	10	3.69	3.2	26	4	0
L Calcarine	PrimVisual (17)	9	3.64	3.17	-14	-80	6

Appendix II.2. Regions showing greater activation for subsequently recognized words than subsequently forgotten words in Figure 3.1

Approx. AAL	Approx. BA	Voxels (kE)	peak T	peak equiv Z	x (mm)	y (mm)	z (mm)
L Supp_Motor_Area	6	1346	8.92	5.68	-4	12	50
L Frontal_Sup_Medial	8		8.46	5.53	-6	24	42
R Cingulum_Mid	8		6.26	4.65	16	12	38
L Insula	Insula (13)	3210	8.28	5.46	-32	20	-2
L Frontal_Inf_Tri	46		7.37	5.13	-44	30	12
L Frontal_Inf_Orb	47		7.17	5.04	-42	18	-8
L Fusiform	Fusiform (37)	1231	7.91	5.33	-40	-60	-12
L Occipital_Inf	19		5.92	4.49	-30	-80	-10
L Calcarine	VisualAssoc (18)		5.9	4.48	-18	-98	-2
R Thalamus	Thalamus (50)	546	6.64	4.82	8	-14	-8
R Pallidum	GlobPal (51)		5.9	4.48	14	2	2
R Thalamus	Thalamus (50)		5.37	4.21	20	-10	2
R Insula	Insula (13)	532	6.35	4.69	32	24	0
R Frontal_Inf_Orb	47		6.18	4.62	30	32	-10
R Insula	Insula (13)		5.15	4.09	40	14	-2
R Fusiform	Fusiform (37)	1301	6.32	4.68	36	-46	-18
R Fusiform	Fusiform (37)		5.85	4.46	34	-38	-24
R Fusiform	VisualAssoc (18)		5.35	4.2	28	-82	-6
R Cingulum_Mid	24	114	6.04	4.55	8	4	30
L Cingulum_Ant	24		5.93	4.49	-4	2	28
L Fusiform	Hippocampus (54)	115	5.98	4.52	-36	-10	-22
L Hippocampus	38		4.81	3.9	-36	0	-26
L Amygdala	34		4	3.41	-32	-2	-14
L Cerebelum_3	Parahip (36)	19	5.91	4.48	-6	-32	-22
L Cerebelum_3	Thalamus (50)	620	5.38	4.22	-4	-24	-14
L Thalamus	Thalamus (50)		5.17	4.11	-16	-6	12
L Pallidum	GlobPal (51)		4.97	4	-16	4	2
R Angular	7	153	5.23	4.14	26	-46	42
R Postcentral	PrimSensory (1)		3.79	3.27	24	-38	42
L Parietal_Sup	7	467	5.21	4.13	-22	-62	48
L Parietal_Sup	7		5.1	4.07	-26	-54	48
L Occipital_Mid	39		4.81	3.91	-26	-66	34
R Vermis_4_5	19	72	4.99	4.01	6	-60	-14
L Cerebelum_6	19		4.16	3.52	-10	-60	-14
L Vermis_9	Fusiform (37)	211	4.97	4	-2	-52	-34
L Vermis_7	VisualAssoc (18)		4.57	3.77	-4	-70	-28
R Cerebelum_9	Fusiform (37)		4.37	3.65	10	-50	-34
L Parietal_Sup	7	13	4.91	3.96	-16	-44	74
R Cerebelum_6	VisualAssoc (18)	74	4.88	3.95	10	-72	-22
L Cerebelum_10	Parahip (36)	82	4.79	3.9	-12	-34	-42
L Vermis_10	Parahip (36)		4.74	3.86	-2	-32	-46
R Cerebelum_10	Fusiform (37)		3.8	3.27	14	-34	-40
R Fusiform	20	25	4.77	3.89	40	-10	-26
R Occipital_Mid	VisualAssoc (18)	88	4.67	3.83	30	-90	18
R Occipital_Mid	19		4.45	3.69	36	-82	16
L Thalamus	Insula (13)	44	4.6	3.78	-28	-24	2

L Pallidum	Putamen (49)		4.28	3.59	-26	-14	0
L Parietal_Inf	40	41	4.46	3.7	-42	-38	42
L Occipital_Mid	19	23	4.42	3.67	-30	-90	16
L Lingual	30	44	4.4	3.66	-18	-44	0
L Supp_Motor_Area	6	7	4.38	3.65	-10	-10	74
R Occipital_Sup	39	67	4.34	3.62	28	-66	44
R Frontal_Inf_Oper	44	56	4.27	3.58	44	10	30
R Frontal_Inf_Oper	44		4.21	3.54	56	12	30
R Occipital_Mid	19	45	4.22	3.55	28	-68	24
R Frontal_Inf_Tri	46	19	4.2	3.54	50	34	14
L Putamen	Putamen (49)	7	4.19	3.53	-34	-2	0
R Frontal_Inf_Tri	45	6	4.12	3.49	42	26	14
L Vermis_4_5	VisualAssoc (18)	26	4.09	3.47	-2	-60	2
L Lingual	VisualAssoc (18)		3.77	3.26	-2	-70	6
L Precentral	PrimMotor (4)	27	4.07	3.46	-34	-18	50
L Postcentral	PrimMotor (4)		3.9	3.34	-42	-16	50
L Temporal_Mid	21	13	4.06	3.45	-60	-34	4
L Cerebelum_Crus1	Fusiform (37)	5	4.06	3.45	-44	-54	-30
L Frontal_Sup	8	8	3.98	3.4	-16	34	48
R Precentral	6	18	3.83	3.3	52	6	40
R Calcarine	PrimVisual (17)	24	3.77	3.26	4	-80	4
L Lingual	VisualAssoc (18)		3.62	3.15	-2	-76	0
R Cingulum_Ant	32	8	3.77	3.26	12	20	26
R Cerebelum_Crus2	VisualAssoc (18)	5	3.73	3.23	8	-76	-38
L Calcarine	PrimVisual (17)	20	3.68	3.19	-10	-74	8
L Calcarine	PrimVisual (17)		3.65	3.18	-12	-84	6
R Calcarine	PrimVisual (17)	5	3.66	3.18	2	-90	10

Appendix II.3. Regions showing significant parametric modulation of imageability in Figure 3.3

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
R Angular	39	306	7.98	5.35	50	-64	36
R Angular	39		5.42	4.24	44	-58	26
L Precuneus	23	1175	7.24	5.07	-6	-54	18
R Precuneus	23		5.4	4.23	10	-52	14
R Precuneus	23		4.54	3.75	6	-54	28
L Fusiform	Fusiform (37)	433	6.79	4.89	-30	-36	-16
L ParaHippocampal	Parahip (36)		6.66	4.83	-28	-38	-8
L Fusiform	Parahip (36)		5.56	4.31	-32	-28	-18
L Cingulum_Ant	24	69	5.87	4.46	-4	26	10
L Cingulum_Ant	24		3.6	3.14	-2	30	20
R Fusiform	Fusiform (37)	107	5.85	4.46	30	-30	-20
R ParaHippocampal	Hippocampus (54)		4.58	3.77	30	-36	-8
L Cingulum_Ant	32	355	5.78	4.42	-12	36	-6
L Olfactory	32		5.69	4.38	-8	26	-10
R Cingulum_Ant	24		4.66	3.82	4	36	-2
L Temporal_Mid	21	77	5.47	4.27	-60	-10	-18
L Temporal_Mid	21		5.09	4.07	-62	-18	-16
L Occipital_Mid	39	745	5.46	4.26	-40	-76	32
L Occipital_Mid	39		5.16	4.1	-30	-78	36
L Angular	39		5.01	4.02	-46	-70	34
R Rectus	11	51	5.14	4.09	4	36	-16
L Frontal_Sup	8	319	5.1	4.07	-20	20	42
L Frontal_Mid	8		5.1	4.07	-22	26	48
L Frontal_Sup	8		4.3	3.6	-18	32	36
L Cingulum_Ant	32	221	4.94	3.98	-2	48	6
R Cingulum_Ant	10		4.92	3.97	6	46	12
L Frontal_Sup_Medial	9		4.1	3.47	-8	44	20
R Temporal_Mid	21	20	4.78	3.89	60	-6	-24
R Frontal_Mid	8	42	4.6	3.78	24	28	42
R Frontal_Sup	8		4.22	3.55	18	36	46
L Frontal_Inf_Orb	47	29	4.27	3.58	-26	34	-10
L Amygdala	Amygdala (53)	9	4.23	3.56	-26	-2	-18
R Frontal_Sup	8	6	4.22	3.55	24	24	60
R Temporal_Mid	21	5	4.19	3.53	60	-14	-16
R SupraMarginal	40	22	4.17	3.52	54	-46	36
L Cingulum_Ant	24	17	4.04	3.43	-4	20	22
R Frontal_Sup_Medial	10	6	3.89	3.34	4	60	16
R Temporal_Pole_Mid	38	5	3.8	3.27	56	10	-28

Appendix II.4. Regions showing significant parametric modulation of phonological distinctiveness in Figure 3.3

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
L Fusiform	19	2660	7.71	5.26	-28	-74	-8
L Cerebelum_6	VisualAssoc (18)		7.41	5.14	-10	-82	-14
R Occipital_Sup	VisualAssoc (18)		7.3	5.1	22	-94	14
R Occipital_Inf	19	228	5.5	4.28	32	-70	-6
R Occipital_Inf	VisualAssoc (18)		5.07	4.05	32	-80	-4
R Angular	39	58	5.06	4.05	36	-54	20
R Calcarine	23		4.18	3.53	28	-56	16
L Fusiform	19	8	4.15	3.51	-36	-50	-2
R Cerebelum_Crus1	Fusiform (37)	10	4.14	3.5	40	-76	-20
L Fusiform	Fusiform (37)	9	4.08	3.46	-32	-62	-14
L Occipital_Sup	7	20	3.87	3.32	-26	-68	24
R Hippocampus	Hippocampus (54)	6	3.85	3.31	36	-12	-16
R Calcarine	PrimVisual (17)	8	3.73	3.23	6	-72	18
L Calcarine	PrimVisual (17)	5	3.67	3.19	-8	-68	14

Appendix II.5. Regions showing significant parametric modulation of semantic distinctiveness in Figure 3.3

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
L Frontal_Inf_Tri	45	156	5.94	4.5	-44	28	10
L Fusiform	Fusiform (37)	227	5.71	4.39	-32	-44	-18
L Fusiform	Fusiform (37)		5.43	4.24	-28	-36	-20
L Fusiform	Fusiform (37)		5.13	4.08	-40	-50	-16
L Occipital_Inf	19	80	5.27	4.16	-42	-68	-6
L Insula	PrimMotor (4)	60	5.24	4.14	-36	-6	18
L Insula	Insula (13)		4.73	3.86	-36	-6	6
L Precentral	6	88	5.15	4.1	-44	2	30
R Vermis_3	Parahip (36)	28	4.75	3.87	0	-32	-6
L Occipital_Mid	7	49	4.6	3.78	-26	-74	32
R Fusiform	Fusiform (37)	7	4.09	3.47	32	-30	-22
R Frontal_Inf_Orb	47	12	3.95	3.38	32	34	-10
R Frontal_Inf_Tri	46	18	3.95	3.38	46	32	12
R Precentral	6	13	3.85	3.31	48	8	32

Appendix II.6. Regions showing significant parametric modulation of judgment reaction time in Figure 3.3

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
L Supp_Motor_Area	8	7765	9.52	5.86	-4	18	48
L Insula	Insula (13)		8.33	5.48	-32	22	0
L Frontal_Inf_Oper	44		7.59	5.21	-42	12	24
R Frontal_Inf_Oper	44	1378	7.24	5.07	58	20	20
R Insula	Insula (13)		6.9	4.93	34	24	2
R Frontal_Inf_Tri	9		6.51	4.76	48	26	22
R Thalamus	Thalamus (50)	532	6.6	4.8	10	-14	-8
R Thalamus	Thalamus (50)		6.43	4.73	4	-16	-2
L Thalamus	Thalamus (50)		6.4	4.72	-4	-20	-8
L Caudate	Caudate (48)	22	5.25	4.15	-4	4	8
L Parietal_Inf	40	538	4.99	4.01	-34	-42	40
L Parietal_Sup	7		4.47	3.71	-20	-62	48
L Parietal_Inf	39		4.46	3.7	-28	-52	42
R Pallidum	GlobPal (51)	26	4.72	3.85	16	0	0
R Vermis_1_2	Parahip (36)	18	4.43	3.68	2	-32	-18
L Pallidum	GlobPal (51)	46	4.39	3.66	-16	-2	0
R Cerebelum_Crus1	VisualAssoc (18)	25	4.39	3.66	10	-74	-28
R Frontal_Inf_Orb	47	60	4.29	3.59	46	38	-14
L Postcentral	PrimSensory (1)	17	4.23	3.56	-48	-32	52
R Cerebelum_6	Fusiform (37)	18	4.12	3.49	24	-58	-28
R Frontal_Mid	6	15	3.77	3.26	34	2	56

Appendix II.7. Regions showing significant parametric modulation of fixed effect prediction of LME (item distinctiveness component) in Figure 3.6

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
L Fusiform	Fusiform (37)	417	8.17	5.42	-30	-36	-18
L Fusiform	Hippocampus (54)		7.25	5.08	-34	-26	-18
L ParaHippocampal	Parahip (36)		3.8	3.27	-20	-32	-12
R Fusiform	Fusiform (37)	142	8.07	5.39	30	-32	-20
R Frontal_Inf_Orb	47	68	5.61	4.34	28	30	-10
L Frontal_Inf_Orb	47	51	4.85	3.93	-28	32	-10
L Temporal_Mid	21	14	4.71	3.85	-62	-10	-18
L Cingulum_Ant	32	15	4.32	3.61	-12	36	-6
L Precuneus	23	57	4.17	3.52	-4	-54	16
L Calcarine	23		3.79	3.27	-10	-52	8
L Rectus	11	7	4.16	3.52	-10	26	-10
L Insula	Insula (13)	6	4.07	3.46	-36	0	-8
L Frontal_Inf_Tri	46	32	4.05	3.44	-40	28	12
R Precuneus	23	6	3.91	3.35	14	-50	14

Appendix II.8. Regions showing significant parametric modulation of random item intercept of LME (mean item memorability component) in Figure 3.7 and Figure 3.8

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
R Frontal_Inf_Orb	47	649	8.97	5.69	30	30	-8
R Frontal_Inf_Tri	45		6.1	4.57	46	32	6
R Frontal_Inf_Orb	45		5.88	4.47	38	30	0
L Frontal_Inf_Orb	47	432	7.66	5.23	-34	32	-8
L Insula	47		6.04	4.55	-30	20	-12
L Insula	Insula (13)		4.72	3.85	-44	8	-6
L ParaHippocampal	Parahip (36)	255	6.72	4.86	-10	-24	-12
R Lingual	Parahip (36)		5.07	4.05	12	-30	-10
L Vermis_3	Parahip (36)		4.93	3.97	-4	-30	-14
L Precuneus	23	108	6.43	4.73	-8	-52	22
L Frontal_Sup_Medial	9	336	6.26	4.65	-6	46	24
L Frontal_Sup_Medial	9		5.42	4.24	-6	42	38
L Frontal_Sup_Medial	10		4.03	3.43	-8	54	18
L Insula	45	40	5.47	4.26	-32	30	6
L Rectus	11	27	4.79	3.9	-6	48	-14
L Cingulum_Ant	32		4.36	3.64	-6	50	-2
R Frontal_Inf_Oper	44	61	4.63	3.8	44	8	28
R Frontal_Inf_Oper	44		3.96	3.39	40	4	22
L Rolandic_Oper	PrimMotor (4)	31	4.52	3.74	-38	-6	18
L Insula	Insula (13)		3.84	3.31	-40	-8	6
L Frontal_Inf_Tri	46	66	4.49	3.72	-44	30	16
R Cerebelum_6	Fusiform (37)	23	4.34	3.62	24	-56	-24
R Temporal_Pole_Sup	38	5	4.32	3.61	40	20	-26
L Amygdala	Amygdala (53)	19	4.24	3.56	-24	-2	-16
L Vermis_9	Fusiform (37)	25	4.22	3.55	0	-54	-32
R Temporal_Sup	Insula (13)	15	4.21	3.54	40	-4	-14
L Temporal_Mid	21	25	4.18	3.53	-58	-10	-16
R Vermis_6	VisualAssoc (18)	40	4.16	3.51	6	-66	-22
L Vermis_6	VisualAssoc (18)		3.72	3.22	-4	-68	-22
R Frontal_Inf_Tri	9	25	4.01	3.42	52	28	18
L Fusiform	Fusiform (37)	16	3.96	3.38	-38	-54	-10
L Cingulum_Ant	24	34	3.96	3.38	-4	6	28
L Cingulum_Ant	24		3.77	3.26	-2	14	26
R Thalamus	Caudate (48)	5	3.95	3.38	16	-22	20
R Cerebelum_4_5	Fusiform (37)	8	3.89	3.34	16	-40	-18
L Cerebelum_4_5	Fusiform (37)	6	3.87	3.33	-30	-34	-26
L Frontal_Sup	8	9	3.87	3.32	-14	30	46
L Cerebelum_6	19	16	3.75	3.25	-28	-62	-26
R Hippocampus	Hippocampus (54)	7	3.67	3.19	26	-38	4

Appendix II.9. Regions showing significant parametric modulation of leave-one-out memorability score in Figure 3.8

Approx. AAL	**Approx. BA**	**Voxels (kE)**	**peak T**	**peak equiv Z**	**x (mm)**	**y (mm)**	**z (mm)**
R Frontal_Inf_Orb	47	570	9.89	5.97	30	30	-8
R Insula	Insula (13)		5.38	4.22	32	20	-14
R Frontal_Inf_Tri	45		5.14	4.09	36	30	4
L Frontal_Inf_Orb	47	493	6.59	4.8	-32	30	-8
L Insula	Insula (13)		6.51	4.77	-30	18	-12
L Insula	Insula (13)		5.35	4.2	-40	4	-10
L ParaHippocampal	Parahip (36)	169	6.55	4.78	-10	-26	-12
R Lingual	Parahip (36)		4.76	3.88	14	-28	-10
R Lingual	Thalamus (50)		4.04	3.44	8	-30	-4
L Frontal_Sup_Medial	9	175	5.71	4.39	-8	44	22
R Cingulum_Ant	10		4.06	3.45	4	52	20
R Frontal_Sup_Medial	9	82	5.33	4.19	0	44	40
L Frontal_Sup_Medial	8		4.68	3.83	-8	40	36
R Temporal_Pole_Mid	38	8	5.18	4.11	44	18	-32
L Precuneus	23	42	5.17	4.11	-8	-52	22
L Rectus	11	10	4.92	3.97	-6	48	-14
R Temporal_Mid	38	20	4.82	3.91	48	2	-30
R Temporal_Inf	38		3.85	3.31	50	10	-34
L Insula	Insula (13)	35	4.41	3.67	-40	-6	4
L Rolandic_Oper	PrimMotor (4)		3.93	3.36	-38	-6	18
R Insula	Insula (13)	22	4.41	3.67	38	-4	-12
L Cingulum_Ant	32	7	4.18	3.52	-6	50	0
L Thalamus	Thalamus (50)	22	4.16	3.52	-6	-6	2
L Cingulum_Ant	24	6	3.96	3.38	-4	26	16
R Temporal_Mid	21	6	3.96	3.38	56	-18	-12
R Frontal_Inf_Oper	44	6	3.94	3.37	44	10	30
R Insula	Insula (13)	7	3.83	3.3	46	4	-4
L Temporal_Mid	21	8	3.73	3.23	-54	-28	-8
L Hippocampus	Insula (13)	5	3.67	3.19	-38	-22	-6
R Cingulum_Mid	24	5	3.66	3.18	2	-6	36

Appendix II.10. Regions showing significant parametric modulation of LME model residual (subject-driven component) in Figure 3.9.

Approx. AAL	Approx. BA	Voxels (kE)	peak T	peak equiv Z	x (mm)	y (mm)	z (mm)
L Supp_Motor_Area	8	1318	9	5.7	-4	20	46
L Supp_Motor_Area	6		8.65	5.59	-2	12	50
L Supp_Motor_Area	6		6.66	4.83	-2	14	60
L Occipital_Mid	VisualAssoc (18)	5890	7.94	5.34	-20	-98	-2
R Calcarine	VisualAssoc (18)		7.55	5.19	22	-96	-2
R Fusiform	Fusiform (37)		7.52	5.18	36	-46	-20
L Frontal_Inf_Tri	45	3283	7.84	5.3	-48	30	6
L Frontal_Inf_Orb	47		7.16	5.04	-32	34	-14
L Insula	47		6.95	4.95	-44	16	-4
R Insula	45	313	7.21	5.06	36	28	0
R Frontal_Inf_Orb	47		5.64	4.35	34	30	-8
R Frontal_Mid_Orb	47		4.48	3.71	36	38	-14
L Parietal_Inf	39	602	6.31	4.67	-26	-66	42
L Parietal_Sup	7		5.46	4.26	-22	-60	50
L Parietal_Sup	7		4.85	3.93	-24	-50	44
R Thalamus	Thalamus (50)	232	5.98	4.52	22	-12	2
R Thalamus	Thalamus (50)		4.97	3.99	18	-18	14
R Thalamus	Thalamus (50)		4.15	3.51	8	-14	-8
L Thalamus	Thalamus (50)	254	5.56	4.31	-14	-20	2
L Thalamus	Thalamus (50)		4.42	3.67	-18	-20	14
L Thalamus	Thalamus (50)		4.04	3.44	-14	-8	12
L Amygdala	Amygdala (53)	116	5.5	4.28	-34	0	-22
L Fusiform	Parahip (36)		5.25	4.15	-38	-12	-26
L Temporal_Mid	38		4.21	3.55	-40	0	-28
R Cingulum_Mid	24	32	5.38	4.22	10	4	30
L Cerebelum_4_5	Thalamus (50)	47	5.35	4.2	-4	-22	-12
R Angular	7	415	5.13	4.08	24	-46	42
R Parietal_Inf	7		4.5	3.72	24	-50	52
R Parietal_Sup	7		4.49	3.72	24	-62	50
L Cingulum_Ant	24	36	4.97	4	-4	2	28
L Precentral	PrimMotor (4)	121	4.96	3.99	-38	-16	50
L Postcentral	PrimSensory (1)		4.2	3.54	-38	-24	42
L Parietal_Inf	40		3.57	3.12	-44	-28	46
R Cerebelum_9	Fusiform (37)	91	4.87	3.94	12	-42	-32
R Cerebelum_9	Fusiform (37)		4.79	3.9	12	-50	-34
L Lingual	VisualAssoc (18)	40	4.85	3.93	-18	-44	-2
L Hippocampus	Putamen (49)	114	4.85	3.93	-28	-24	-2
L Putamen	Putamen (49)		4.53	3.74	-28	-14	-2
L Putamen	Putamen (49)		4.14	3.5	-34	-4	0

R Insula	44	83	4.78	3.89	48	8	4
R Insula	44		4.37	3.64	42	14	2
R Insula	Insula (13)		3.8	3.28	34	8	4
R Vermis_10	Parahip (36)	50	4.67	3.83	4	-32	-40
L Vermis_10	Parahip (36)		3.95	3.38	-2	-32	-46
R Cerebelum_10	Fusiform (37)		3.92	3.36	12	-34	-40
L Lingual	VisualAssoc (18)	27	4.36	3.64	-2	-66	6
L Insula	PrimSensory (1)	11	4.33	3.62	-38	-22	28
L Parietal_Inf	PrimSensory (1)	35	4.24	3.57	-40	-36	42
R Frontal_Sup	6	7	4.24	3.56	14	-12	74
R Frontal_Inf_Tri	45	8	4.24	3.56	40	24	14
L Parietal_Sup	7	9	4.21	3.54	-16	-44	74
L Frontal_Sup	8	10	4.17	3.52	-16	34	48
L Cerebelum_10	Parahip (36)	14	4.16	3.52	-12	-34	-42
R Pallidum	GlobPal (51)	22	4.15	3.51	14	4	2
L Temporal_Mid	21	12	4.05	3.44	-58	-36	6
L Insula	Insula (13)	7	4	3.41	-34	8	0
L Cerebelum_10	Parahip (36)	5	4	3.41	-12	-28	-34
R Heschl	PrimAuditory (41)	8	3.99	3.4	30	-34	16
R Precentral	PrimMotor (4)	6	3.97	3.39	34	-14	32
R Precentral	6	8	3.95	3.38	58	-2	42
L Calcarine	PrimVisual (17)	29	3.94	3.37	-16	-80	10
L Lingual	PrimVisual (17)		3.59	3.13	-10	-74	4
L Hippocampus	Putamen (49)	7	3.94	3.37	-22	-16	-6
R Cingulum_Mid	8	12	3.93	3.36	14	24	30
R Calcarine	PrimVisual (17)	14	3.83	3.3	8	-84	6
L Rectus	11	5	3.82	3.29	-4	40	-16
R Supp_Motor_Area	6	5	3.79	3.27	8	0	72
L Supp_Motor_Area	6	5	3.78	3.26	-10	-10	74
R Calcarine	PrimVisual (17)	5	3.71	3.22	0	-74	14
L Postcentral	PrimSensory (1)	7	3.67	3.19	-42	-32	54
L Temporal_Mid	21	6	3.62	3.15	-52	-46	2

Appendix III. NAART and NAART 35 sample scoring sheet

Only the items in bold (NAART 35) were adopted for this study.

NAART

Sample Scoring Sheet

Page 1

DEBT det
DEBRIS də·brē, dā·brē′, dā′·brē
AISLE īl
REIGN rān
DEPOT dē,·pō, de′·pō
SIMILE sim′·ə·lē
LINGERIE lan′·zhə·rē′, lon′·zhə·rā′
RECIPE res′·ə·pē
GOUGE gauj
HEIR ãr
SUBTLE sət′·əl
CATACOMB kat′·ə·kōm
BOUQUET bō·kā′, bü·kā′
GAUGE gāj
COLONEL kərn′·əl

SUBPOENA se·pē′·nə
PLACEBO plə·sē′·bō
PROCREATE prō′·krē·āt
PSALM säm, sälm*
BANAL bə·nȧl′, bā·nal′, bān′·əl
RAREFY rãr′·ə·fī
GIST jist
CORPS kor, korz
HORS D'OEUVRE ȯr′ dərv(r)′
SIEVE siv
HIATUS hī·ā·təs
GAUCHE gōsh
ZEALOT zel′·ət
PARADIGM par′·ə·dīm, par′·ə·dim
FACADE fə·säd′

Page 2

CELLIST chel′·əst
INDICT in·dīt′
DETENTE dā·tä(n)t
IMPUGN im·pyün′
CAPON kā′·pən, kā′·pon
RADIX rād′·iks
AEON ē′·ən, e′·an
EPITOME i·pit′·ə·mē
EQUIVOCAL i·kwiv′·ə·kəl
REIFY rā′·ə·fī, rē′·ə·fi
INDICES in′·də·sēz
ASSIGNATE as′·ig·nāt′
TOPIARY tō·pē·er′·ē
CAVEAT kav′·ē·at, kāv′·ē·at, kā·vē·at′**
SUPERFLUOUS sü·pėr′·flü·əs

LEVIATHAN li·vī′·ə·thən
PRELATE prel′·ət, prē′·lāt*
QUADRUPED kwäd′·rə·ped
SIDEREAL sī·dir′·ē·al, sə·dir′·ē·al
ABSTEMIOUS ab·stē′·mē·əs
BEATIFY bē·at′·ə·fī
GAOLED jāld
DEMESNE di·mān′, di·mēn′
SYNCOPE sing′·kə·pē, sin′·k′rrn·pē
ENNUI an·wē′
DRACHM dram
CIDEVANT sēd·ə·vä(n)′
EPERGNE i·pərn′, ā·pərn′
VIVACE vē·väch′·ā, vē·väch′·ē
TALIPES tal′·ə·pēz
SYNECDOCHE sə·nek′·də·kē

References

Addis, D. R., Moscovitch, M., Crawley, A. P., & McAndrews, M. P. (2004). Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus*, *14*(6), 752–762. https://doi.org/10.1002/hipo.10215

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Aggleton, J., & Brown, M. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and Brain Sciences*, *22*(03), 425–444. Retrieved from http://journals.cambridge.org/abstract_S0140525X99002034

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, *17*(8), 379–390. https://doi.org/10.1016/j.tics.2013.06.009

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bainbridge, W. A. (2016). The Memorability of People: Intrinsic Memorability Across Transformations of a Person{\textquoteright}s Face. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1–12. https://doi.org/10.1037/xlm0000339

Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, *149*, 141–152. https://doi.org/10.1016/j.neuroimage.2017.01.063

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology. General*, *142*(4), 1323–1334. https://doi.org/10.1037/a0033872

Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports*, 1–11. https://doi.org/10.1038/s41598-018-26467-5

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The Word-Frequency Mirror Effect in Young, Old, and Early-Stage Alzheimer's Disease: Evidence for Two Processes in Episodic Recognition Performance. *Journal of Memory and Language*, *46*(1), 199–226. https://doi.org/10.1006/jmla.2001.2803

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(5), 576–587. https://doi.org/10.1037/0278-7393.6.5.576

Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes Unseen: The Parahippocampal Cortex Intrinsically Subserves Contextual Associations, Not Scenes or Places Per Se. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.0987-08.2008

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random e ff ects structure for confirmatory hypothesis testing : Keep it maximal. *Journal of Memory and Language*, *44*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=16021798&retmode=ref&cmd=prlinks

Bonner, M. F., Price, A. R., Peelle, J. E., & Grossman, M. (2016). Semantics of the Visual Environment

Encoded in Parahippocampal Cortex. *Journal of Cognitive Neuroscience*, *28*(3), 361–378. https://doi.org/10.1162/jocn_a_00908
Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*(12), 539–546. https://doi.org/10.1016/j.tics.2004.10.003
Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of …*, *29*(3), 461–473. https://doi.org/10.1080/14640747708400622
Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*(1), 51–61. https://doi.org/10.1038/35049064
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977
Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5
Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*. https://doi.org/10.1196/annals.1440.011
Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004
Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, *116*(Part B), 165–178. https://doi.org/10.1016/j.visres.2015.03.005
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, M. W. L., Westphal, C., Goh, J., Graham, S., & Song, A. W. (2003). Word frequency and subsequent memory effects studied using event-related fMRI. *NeuroImage*, *20*(2), 1042–1051. https://doi.org/10.1016/S1053-8119(03)00335-5
Clark, D., & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia*, *41*(3), 304–317. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12457756&retmode=ref&cmd=prlinks
Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/s0022-5371(73)80014-3
Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*. https://doi.org/10.1177/014662168300700301
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
Cortese, M. J., Khanna, M. M., & Hacker, S. (2010). Recognition memory for 2,578 monosyllabic words. *Memory*, *18*(6), 595–609.
Cortese, M. J., McCarty, D. P., & Schock, J. (2015). A mega recognition memory study of 2897 disyllabic words. *Quarterly Journal of Experimental Psychology (2006)*, *68*(8), 1489–1501.
Cortese, M. J., Watson, J. M., Wang, J., & Fugett, A. (2004). Relating distinctive orthographic and phonological processes to episodic memory performance. *Memory & Cognition*, *32*(4), 632–639. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=15478757&retmode=ref&cmd=prlinks
Cox, G. E., Hemmer, P., Aue, W. R., & Criss, A. H. (2018). Information and processes underlying

semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General*, *147*(4), 545–590. https://doi.org/10.1037/xge0000407
Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x
Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology. General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268
Criss, A. H., & Malmberg, K. J. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language*, *59*(3), 331–345. https://doi.org/10.1016/j.jml.2008.05.002
Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, *5*(5), 329–340. https://doi.org/10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.0.CO;2-5
Daselaar, S. M., Veltman, D. J., Rombouts, S. A. R. B., Raaijmakers, J. G. W., & Jonker, C. (2003). Neuroanatomical correlates of episodic encoding and retrieval in young and elderly subjects. *Brain*, *126*(1), 43–56. https://doi.org/10.1093/brain/awg005
Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*(6), 693–700. https://doi.org/10.1016/j.conb.2006.10.012
Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple Routes to Memory: Distinct Medial Temporal Lobe Processes Build Item and Source Memories. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(4), 2157–2162. https://doi.org/10.2307/3138501?refreqid=search-gateway:4aa206577611127ea20bdef7c262ef80
Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341. https://doi.org/10.1016/j.tics.2005.05.004
Dehaene, S., Gurvan, C. A., Clec, L., Poline, J., Bihan, D. Le, & Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport. *Neuroreport*, *13*(3). https://doi.org/10.1097/00001756-200203040-00015
Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00303.2011
Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2007.08.001
Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. *Hippocampus*, *18*(6), 536–541. https://doi.org/10.1002/hipo.20433
Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2013). Parahippocampal cortex activation during context reinstatement predicts item recollection. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/a0034029
Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *NeuroImage*, *23*(1), 64–74. https://doi.org/10.1016/j.neuroimage.2004.05.015
Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The Medial Temporal Lobe and Recognition Memory. *Annual Review of Neuroscience*, *30*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328
Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory*. https://doi.org/10.1037/0278-7393.6.5.588
Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2008.07.004

Fletcher, P. C., Frith, C. D., Baker, S. C., Shallice, T., Frackowiak, R. S., & Dolan, R. J. (1995). The mind's eye--precuneus activation in memory-related imagery. *NeuroImage*, *2*(3), 195–200. https://doi.org/10.1006/nimg.1995.1025

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *NeuroImage*, *32*(3), 1413–1421. https://doi.org/10.1016/j.neuroimage.2006.06.007

Freeman, E., Heathcote, A., Chalmers, K., & Hockley, W. (2010). Item effects in recognition memory for words. *Journal of Memory and Language*, *62*(1), 1–18. https://doi.org/10.1016/j.jml.2009.09.004

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012

Glanc, G. A., & Greene, R. L. (2007). Orthographic neighborhood size effects in recognition memory. *Memory & Cognition*, *35*(2), 365–371. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=17645177&retmode=ref&cmd=prlinks

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=4010518&retmode=ref&cmd=prlinks

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2002). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.071043098

Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using Imagination to Understand the Neural Basis of Episodic Memory. *Journal of Neuroscience*, *27*(52), 14365–14374. https://doi.org/10.1523/jneurosci.4549-07.2007

Hassabis, Demis, Chu, C., Rees, G., Weiskopf, N., Molyneux, P. D., & Maguire, E. A. (2009). Decoding Neuronal Ensembles in the Human Hippocampus. *Current Biology*. https://doi.org/10.1016/j.cub.2009.02.033

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. https://doi.org/10.1037/a0036306

Hemmer, P., & Criss, A. H. (2013). The shape of things to come: evaluating word frequency as a continuous variable in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1947–1952. https://doi.org/10.1037/a0033744

Hirshman, E., & Jackson, E. (1997). Distinctive perceptual processing and memory. *Journal of Memory and Language*, *36*(1), 2–12. https://doi.org/10.1006/jmla.1996.2470

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

Isoda, M., & Noritake, A. (2013). What makes the dorsomedial frontal cortex active during reading the mental states of others? *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2013.00232

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America*, *132*(2), EL74--EL80. https://doi.org/10.1121/1.4731641

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The Role of Semantic Diversity in Word Recognition across Aging and Bilingualism. *Frontiers in Psychology*, *7*(250), 811–814. https://doi.org/10.3389/fpsyg.2016.00703

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, *66*(2), 115–124. Retrieved from http://psycnet.apa.org/record/2012-15094-006

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004).

Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023–1026. https://doi.org/10.1126/science.1089910
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
Kirchhoff, B. A., Schapiro, M. L., & Buckner, R. L. (2005). Orthographic distinctiveness and semantic elaboration provide separate contributions to memory. *Journal of Cognitive Neuroscience*, *17*(12), 1841–1854. https://doi.org/10.1162/089892905775008670
Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *20*(16), 6173–6180. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10934267&retmode=ref&cmd=prlinks
Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, *23*(4), 425–449. https://doi.org/10.1016/s0022-5371(84)90282-2
Koriat, A. (2008). Subjective Confidence in One's Answers: The Consensuality Principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 945–959. https://doi.org/10.1037/0278-7393.34.4.945
Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *NeuroImage*, *21*(3), 946–953. https://doi.org/10.1016/j.neuroimage.2003.10.021
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211
LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(13), 5466–5474. https://doi.org/10.1523/JNEUROSCI.4293-12.2013
Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *The Quarterly Journal of Experimental Psychology*, *36*, 174702181773983--16. https://doi.org/10.1177/1747021817739834
Liang, J. C., Wagner, A. D., & Preston, A. R. (2012). Content Representation in the Human Medial Temporal Lobe. *Cerebral Cortex*, *23*(1), 80–96. https://doi.org/10.1093/cercor/bhr379
Long, D. L., & Prat, C. S. (2002). Memory for Star Trek : The role of prior knowledge in recognition revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1073–1082. https://doi.org/10.1037//0278-7393.28.6.1073
Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). *Learning word vectors for sentiment analysis*. Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2002472.2002491
Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *356*(1413), 1441–1451. https://doi.org/10.1098/rstb.2001.0944
Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12699141&retmode=ref&cmd=prlinks
Mandera Pawełand Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001
McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Opinion TRENDS in Cognitive Sciences*, *7*(7), 293.

https://doi.org/10.1016/S1364-6613(03)00134-7
Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/s0022-5371(77)80016-9
Mumford, J. A., Poline, J.-B., & Poldrack, R. A. (2015). Orthogonalization of Regressors in fMRI Models. *PLoS ONE*, *10*(4), e0126255--11. https://doi.org/10.1371/journal.pone.0126255
Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive Memory : Survival Processing Enhances Retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 263–273. https://doi.org/10.1037/0278-7393.33.2.263
Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the Living: Episodic Memory is tuned to animacy . *Current Directions in Psychological Science*, *26*(1), 22–27. https://doi.org/10.1177/0963721416667711
Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: the mnemonic value of animacy. *Psychological Science*, *24*(10), 2099–2105. https://doi.org/10.1177/0956797613480803
Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, *14*(134), 20170211–20170213. https://doi.org/10.1098/rsif.2017.0213
Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, Imagery, and Meaningfulness Values for 925 Nouns. *Journal of Experimental Psychology*, *76*(1 PART 2), 1–25. https://doi.org/10.1037/h0025327
Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*, *67*(4), 360–371. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=2441971&retmode=ref&cmd=prlinks
Park, H., Reder, L. M., & Dickison, D. (2005). The Effects of Word Frequency and Similarity on Recognition Judgments: The Role of Recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 568–578. https://doi.org/10.1037/0278-7393.31.3.568
Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). The Neural Consequences of Semantic Richness: When More Comes to Mind, Less Activation Is Observed. *Psychological Science*, *18*(5), 401–406. https://doi.org/10.2307/40064630?refreqid=search-gateway:5c81d703dd0f8df067a51b2cd41f4ebc
Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., … Petersen, S. E. (2011). Functional Network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006
Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, *19*(3), 473–481. https://doi.org/10.1016/S1053-8119(03)00084-3
Qin, S., van Marle, H. J. F., Hermans, E. J., & Fernández, G. (2011). Subjective sense of memory strength and the objective amount of information accurately remembered are related to distinct neural correlates at encoding. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(24), 8920–8927. https://doi.org/10.1523/JNEUROSCI.2587-10.2011
Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.98.2.676
Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology. General*, *140*(3), 464–487. https://doi.org/10.1037/a0023810
Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*(2), 752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035
Roediger III, H. L., & Gallo, D. A. (2001). Levels of processing: Some unanswered questions.

...cognitive Aging: Essays in Honour of .... Retrieved from http://books.google.com/books?hl=en&lr=&id=Yh5bTJ7T5fIC&oi=fnd&pg=PA28&dq=Levels+of+Processing+Some+Unanswered+Questions+Roediger+and+Gallo&ots=GEfOZLBW9P&sig=N2yBEw0Oe7PilX9DL39rTnmB8PQ

Rohde, D. L. T., Gonnerman, L. M., & Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM*.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, *27*(1), 188–200. https://doi.org/10.1016/j.neuroimage.2005.04.012

Schock, J., Cortese, M. J., Khanna, M. M., & Toppi, S. (2012). Age of acquisition estimates for 3,000 disyllabic words. *Behavior Research Methods*, *44*(4), 971–977. https://doi.org/10.3758/s13428-012-0209-x

Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., & Richlan, F. (2016). Words in Context: The Effects of Length, Frequency, and Predictability on Brain Responses During Natural Reading. *Cerebral Cortex*, *26*(10), 3889.2--3904. https://doi.org/10.1093/cercor/bhw184

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=1896849&retmode=ref&cmd=prlinks

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia*, *44*(12), 2189–2208. https://doi.org/10.1016/j.neuropsychologia.2006.05.023

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology: ...*, *86*(2), 255–262. https://doi.org/10.1037/h0029997

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Turner, B. O., Mumford, J. A., Poldrack, R. A., & Ashby, F. G. (2012). Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. *NeuroImage*, *62*(3), 1429–1438. https://doi.org/10.1016/j.neuroimage.2012.05.057

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. https://doi.org/10.1613/jair.2934

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of Learning and Memory*, *91*(2), 139–154. https://doi.org/10.1016/j.nlm.2008.10.011

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*, *22*(3), 537–549. https://doi.org/10.1093/cercor/bhr100

Vogel, A. C., Petersen, S. E., & Schlaggar, B. L. (2014). The VWFA: it's not just for words anymore. *Frontiers in Human Neuroscience*, *8*, 88. https://doi.org/10.3389/fnhum.2014.00088

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., ... Buckner, R. L. (1998). Building Memories: Remembering and Forgetting of Verbal Experiences as Predicted by Brain Activity. *Science*, *281*(5380), 1188–1191. https://doi.org/10.2307/2895504?ref=search-gateway:2eb55970b2285ffdeab558d213f9774f

Westfall, J., Nichols, T. E., & Yarkoni, T. (2016). Fixing the stimulus-as-fixed-effect fallacy in task fMRI. *Wellcome Open Research*, *1*, 23. https://doi.org/10.12688/wellcomeopenres.10298.1

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin and Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Yarkoni, T., Speer, N. K., Balota, D. A., McAvoy, M. P., & Zacks, J. M. (2008). Pictures of a thousand words: Investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage*, *42*(2), 973–987. https://doi.org/10.1016/j.neuroimage.2008.04.258

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011