ON THE MEANING OF NUMBERS: FLEXIBILITY IN THE STRUCTURE AND RETRIEVAL OF MEMORIES FOR ARABIC NUMERALS

BY

DANIELLE DICKSON

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Kara Federmeier, Chair
Professor Monica Fabiani
Professor Charles Hillman
Professor Cynthia Fisher
Assistant Professor Daniel Hyde

Abstract

We read symbolic representations of numbers like “24” across a multitude of contexts – as the name of a TV show, the answer to common arithmetic problems, a symbol for the linguistic expression “twenty-four”, among others – and utilize multiple systems of memory in order to appropriately interpret them. This thesis examines how these meanings of Arabic numerals are flexibly accessed, retrieved, and evaluated by healthy college-age adults. In order to dissociate these rapidly occurring processes, event-related potentials (ERPs) were recorded while participants read common numerals in tasks that differed in the type and amount of numeral-associated information that would need to be recalled.

The first two experiments specifically looked at the evaluation of Arabic numerals in arithmetic contexts, and examined how the two cerebral hemispheres approach reading equations and evaluating potential answers. These experiments revealed similarities in how the hemispheres respond to contextually congruous and incongruous answers but differences in how they evaluate other aspects of provided answers. Specifically, the right hemisphere (and not the left) is sensitive to mathematical relationships beyond whether an answer is right or wrong.

The second two experiments assessed how relatively more automatic access of meaning during numeral reading is influenced by task goals (Experiment 3) or by item-level properties of numerals (Experiment 4). The results showed that the amount of meaning information that is relatively *automatically* accessed during numeral reading is similar (and small) across task, but that the information that can be *deliberately or explicitly* retrieved differs across item type depending on personalized ratings of familiarity. Additionally, the nature of what is automatically retrieved from semantics is at least somewhat malleable, because, whereas

Experiment 4 obtained effects similar in important ways to those observed during semantic retrieval for words, Experiment 3 did not.

Across all experiments, the results speak to a fluidity in the kind of information that can be brought to bear during numeral processing, depending on what sort of contextual support is provided and which types of evaluative processes are needed in order to perform the task at hand.

ACKNOWLEDGMENTS

I would like to thank Professor Kara D. Federmeier for supporting my extensive adventures in studying arithmetic and number processing. She was the best possible advisor for my research program, providing thoughtful insight and guidance at every stage of development. I would also like to thank the other members of my dissertation committee, Monica Fabiani, Charles Hillman, Cynthia Fisher and Daniel Hyde, for their thoughtful and engaging feedback on my thesis and current research directions. I have also more generally been fortunate to work in a vibrant laboratory culture, headed by Kara, with colleagues who are excited to think about how our different lines of research mutually inform one another. Their collective support through my graduate school career not only improved my research but also made the process fun.

TABLE OF CONTENTS

CHAPTER 1: **INTRODUCTION**……………………………………………………………………………1

CHAPTER 2: **EXPERIMENT 1: HEMISPHERIC DIFFERENCES IN ARITHMETIC FACT EVALUATION AS REVEALED BY ERPS** ……………………………………………………………25

CHAPTER 3: **EXPERIMENT 2: BEHAVIORAL EFFECTS OF ANSWER EVALUATION ACROSS THE HEMISPHERES**……………………………………………………………………………64

CHAPTER 4: **EXPERIMENT 3: ON THE POTENTIAL (IN)FLEXIBILITY OF SEMANTIC MEMORY ACCESS FOR NUMERALS: THE EFFECT OF INCIDENTAL REPETITION ACROSS TASKS** ……………………………………………………………………82

CHAPTER 5: **EXPERIMENT 4: IS YOUR FAVORITE NUMBER SPECIAL? EVIDENCE FOR ITEM-LEVEL DIFFERENCES IN RETRIEVAL OF INFORMATION FROM NUMERALS**……………………………………………………………………………………109

CHAPTER 6: **CONCLUSION**……………………………………………………………………………136

REFERENCES…………………………………………………………………………………………142

CHAPTER 1

Introduction

Numerical cognition: basic number systems

What is the nature of the representational space for numerical knowledge? In the domain of numerical cognition, one of the largest theoretical distinctions is between a *preverbal* number system, which appears to be shared by humans and animals, and a *verbal* number system, which must be deliberately learned and allows for manipulation of numbers in more precise quantities (i.e., the system used during arithmetic calculation, to figure out how much to tip in a restaurant, etc). The common view of the preverbal system is that there is a distinction between small numerosities (1-3), which can be individually subitized accurately, and larger numerosities, which can be distinguished without using verbal counting skills only once the ratio between compared numerosities is sufficiently high (see, e.g., reviews by Brannon & Roitman, 2003; Spelke & Dehaene, 1999; and more recently by Gaber & Schlimm, 2015).

The (learned) verbal system is supported both by number words, used in counting, and, in written language, it is supported by symbols that are used as visual representations of numbers (called numerals), the most commonly used of these symbol systems being Arabic numerals (although Roman numerals function similarly). Numerals exist in an odd middle-ground for educated adults: they are not technically word forms, because their representations do not use letters that map onto phonemes, but they are also not quite pictorial representations either, both because their mapping to meaning is arbitrary and because there is no ambiguity in how their representation should be interpreted (i.e., a picture of a dog could indicate a specific dog, a generic dog, a living animal, etc., whereas the Arabic numeral "1" always maps onto the word "one"). There is evidence that these numerals are increasingly independent of their corresponding verbal (word) representations as adults become more familiar with them, most notably in cases of preserved calculation abilities in patients with various forms of aphasia (Anderson, Damasio, & Damasio, 1990; Cipolotti, Warrington, & Butterworth, 1995; Rossor, Warrington & Cipolotti, 1995; Cipolotti, 1995).

Numerical cognition: representation of learned numerical information in long-term memory

Descriptions of the cognitive representation of numerical information often invoke a number line analogy, with smaller numbers on the left and increasingly larger numbers to the right. Evidence in support of this idea comes from the SNARC (Spatial-Numerical Association of Response Codes) effect, which is found when participants compare pairs of quantities. If smaller quantities are on the right (incompatible with the hypothesized number line), reaction times are slower and accuracies are reduced relative to compatible arrangements of number pairs (where smaller quantities are on the left). This suggests there is a spatial structure to numerical (ordinal) information (Dehaene, Bossini & Giraux, 1993; Fias, Brysbaert, Geypens, & d'Ydewalle, 1996; reviewed in Fischer & Fias, 2005). However, there appears to be no evidence for this linear organization in the neurons of the brain, leading to suggestions that it is a strategy people invoke to lessen working memory loads (Núñez, 2011; Nieder, Freedman, & Miller, 2002; reviewed in Gaber & Schlimm, 2015).

When participants judge and compare quantities of numbers, solve arithmetic problems, or assess correctness of provided answers, problems involving larger numbers are both slower and less accurate (e.g, Campbell & Graham, 1985). Explanations for this "number size" effect have varied in the decades following its discovery, with different theories implicating memory itself, retrieval processes, and/or differences in strategy use across number sizes (Zbrodoff & Logan, 2005, review the chronological unfolding of theories in this domain). The suggestion most relevant for characterizing the cognitive structure of numerical representations has been that increasing number size has a negative effect on the precision of numbers in representational space (Dehaene & Changeux, 1993; Gallistel & Gelman, 1992; Zorzi & Butterworth, 1999; Cohen Kadosh, Tzelgov, Henik, 2008).

Larger numbers, then, are represented less well (or are less distinct in long-term memory) when compared to the more precise representation of smaller numbers. One causal explanation for this difference is that smaller numbers are encountered more frequently than larger numbers. Support for this notion comes from corpus findings that arithmetic problems involving smaller digits are more often used in textbooks than are problems with larger digits (Hamman & Ashcraft, 1986; Ashcraft & Christy, 1995).

When larger numbers are assessed, it is unclear to what extent the magnitude of symbolic numbers is represented holistically (“46” as a unit, Dehaene, Dupoux, & Mehler, 1990), with more continuous processing along a single number line (Reynvoet & Brysbaert, 1999), or whether the processing of these symbols involves decomposition into the tens place and a unit place (“46” as “4” and “6”) (McCloskey, 1992; Nuerk et al., 2001).

Comparison: Long-term semantic memory representation for words

Much like the longstanding interest in how numerical knowledge is represented in long-term memory, in the language domain, understanding the specifics of the structure of the long-term semantic memory system for *words* has also been a topic of interest historically. Much of the most recent progress in this domain over the past two decades has taken advantage of temporally sensitive, functionally-linked event-related potential (ERPs) components to better characterize semantic memory (reviewed in Kutas & Federmeier, 2011). A review of the methods that have been employed in this domain follows, in order to serve as a backdrop for how similar methods can be applied to symbolic number representations.

N400 effects in language comprehension: general

One fruitful approach for examining the structure of meaningful information in long-term memory has been by taking advantage of the response properties of the N400 component. The N400, so named for the timing of its negative-going peak just prior to 400ms post-stimulus onset, is elicited by potentially meaningful stimuli in any sensory modality, and it is temporally stable, appearing reliably in the same timewindow across stimulus modality and task contexts (Kutas & Federmeier, 2011). Its amplitude (amount of negativity at the peak) is related to the amount of *new* semantic information that must be accessed when presented with a given input. The simplest example of this is in comparisons between words (or word-like pseudowords, e.g., “LULP”) and jumbles of letters that do not resemble real words (“TXMR”). The N400 is larger in response to words compared to these character strings because words, when they are presented out of context, activate a lot of new information in semantic memory (especially when they resemble many other real words), whereas vowel-less character strings fail to make connections with as rich a semantic network (Rugg & Nagy, 1987; Holcomb, Grainger & ORourke, 2002; Laszlo & Federmeier, 2011; Laszlo & Federmeier, 2014). This relatively simple finding elegantly revealed one feature of the structure of semantic memory for written words – general orthographic structure of text, at least by 400ms, is more influential than the familiarity of the individual item.

This effect of orthography also reveals that N400 amplitudes reflect a coarser contact with semantic memory than is sometimes assumed to be the case. Given no prior context (i.e., if items are presented in an unstructured list), the N400 amplitude to an individual item does not necessarily reflect how meaningful that item itself is in long-term memory. That interpretation of N400 amplitude would predict that the *meaningful* acronym "DVD" should elicit a larger N400 than a *nonmeaningful* but orthographically regular pseudoword like "DAWK" – in fact, the opposite is true. This is because N400 amplitude to a stimulus reflects an initial, messy, contact with semantic memory (which begins around 250ms post-stimulus presentation and peaks at 400ms), and does not reflect what is ultimately retrieved from memory. Thus, during the first brush with semantic memory, an orthographically regular non-word makes contact with real words that look similar to it, eliciting a larger N400 than orthographically irregular items that have less confusability with neighboring items. In unstructured list contexts, the more delicate process of recognizing "DVD" as a specific item with which a viewer has experience is reflected on components other than the N400 (discussed later).

However, within the class of orthographically regular words, frequency can also influence N400 amplitude: less frequent words result in higher amplitude N400s than do words that are higher in frequency (Smith & Halgren, 1986; Van Petten & Kutas, 1990; Rugg, 1990). There are at least two explanations for this finding, neither of which are particularly in conflict. Either low frequency words require more *new* semantic information to be activated (thus resulting in larger N400s for low frequency words), or high frequency words are more facilitated because of how regularly their semantics are retrieved (thus resulting in *smaller* N400s for high frequency words). This effect is sensitive to task effects – in cases where a particular semantic subcategory is operating as a target, frequency effects are attenuated, whereas cases where the task is more neutral with respect to semantic content (e.g., having to read the words out loud), frequency effects are present (Fischer-Baum, Dickson, & Federmeier, 2014). Tasks themselves can therefore act as a context that influence N400 amplitude of a given stimulus.

Contextual effects on N400: repetition effects

More nuanced properties of semantic memory retrieval during word processing often rely on repetition effects to make inferences. Repetition designs, especially when there are multiple intervening items between repetitions, can be viewed similarly to tests of implicit memory. Specifically, the size of the reduction in N400 amplitude upon an item's repetition can be interpreted as an indication of the extent to which the item was (recently) processed in semantic memory (i.e., more reduction implies more substantial semantic memory processing occurred). Put another way, when an item is still active in semantic memory, less new information about it must be brought to mind than when it was first presented to the system, and the N400 amplitude is therefore smaller on second presentation compared to first presentation.

It was previously described that the amplitude on the N400 on initial presentation of wordforms in lists differentiates between words and word-like pseudowords (large N400 amplitudes) and illegal strings of characters (smaller N400s) (Rugg & Nagy, 1987). When these items are repeated, both words and pseudowords are facilitated (smaller N400s) compared to their initial presentation – there is less *new* semantic information that needs to be accessed when they are seen a second time because of the processing that already occurred.

Initial research suggested that strings of illegal characters did not have access to semantics because they did not show effects of repetition (Rugg & Nagy, 1987). However, it is the case that some strings of unpronounceable letters form an exception: acronyms. Meaningful acronyms like "DVD", which otherwise have small overall N400 amplitudes during first presentation (because they do not resemble real words), *do* show effects of repetition (Laszlo & Federmeier, 2007). The interpretation of this finding was that, unlike non-meaningful strings of letters, acronyms *are* able to make contact with existing representations in long-term memory, which promotes their retention in memory, and leads to their facilitation when they are presented a second time. A similar finding exists with visual items that participants do not have pre-experimental exposure to: novel geometric shapes ("squiggles"), which sometimes might appear to be resemble a familiar object or creature. Shapes were categorized as familiar or unfamiliar, and repetition effects for each were measured within participants on the basis of their personal rating. When these effects were averaged across participants, it was found that only shapes that had been rated familiar (i.e., those which contacted existing semantic memory representations) elicited an N400 repetition effect (Voss, Schendan, & Paller, 2010).

Magnitude of repetition effects differ by length of lag, with the effect reducing in magnitude due to more time passing or due to more intervening items, as well as due to the depth of encoding required by the task, with effects being attenuated for a "shallow" task (just requiring detection of font size differences) compared to larger effects present for "deeper" encoding tasks like detection of particular semantic targets, or tasks that render otherwise forgettable items more relevant (Nagy & Rugg, 1989; Rugg, Furda, & Lorist, 1988; reviewed in Rugg & Doyle, 1994; Laszlo, Stites & Federmeier, 2012; cf, Bentin & Peled, 1990).

In sum, the existence and magnitude of N400 repetition effects have a history of being used to make inferences about the nature of representations that are accessed in long-term memory and how well that information is retained over time. Such methods have never been systematically applied to numeral processing (e.g., is there a repetition effect for common Arabic numerals like “24”?), but it seems likely that this relatively simple approach can also be informative for learning about the nature of how symbolic numbers are processed.

Numerical cognition: arithmetic problem solving

Educated adults can do more with Arabic numerals than assess individual digits for properties like magnitude – they can solve arithmetic problems, retrieving solutions through a series of operations or directly retrieving overlearned answers from long-term memory. Consistent with modern theories of semantic information in language representations, there is a (fairly dominant) model that during arithmetic fact retrieval, numerical information is accessed through spreading activations across a network of inter-related facts about arithmetic (Campbell & Graham, 1985; Campbell, 1987a; see reviews in Fayol & Seron, 2005; Brysbaet, 2005).

At a behavioral level, this spreading of information leads to delayed RTs in verification tasks wherein subjects must respond to provided answers with a confirmation/rejection of the provided answer. In initial work using this task (instead of the production tasks that had dominated the literature), Campbell (1987b) observed delays in RT when participants needed to reject an answer as incorrect if that answer shared properties with the (not presented) correct answer. These associations between answer types in long-term memory are now understood to include confusions at the level of operation (e.g., delayed responses for 2 x 4 = 6, because 6 is the correct answer for the operation of addition; Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986), confusions at the level of digit (e.g., errors are typically consistent with the decade of the correct answer; Domahs, Delazer & Nuerk, 2004), and, finally, so-called table-related confusions in multiplication (e.g., 6 x 5 = 36 is difficult to reject because “36” is the correct answer for a different multiple of 6) (Stazyk, Ashcraft & Hamann, 1982; Campbell, 1987a,b).

This evidence is used to argue in support of a view of memory for symbolic numbers that is like a highly structured network, in contrast to earlier views that answer retrieval and verification was independent of other known arithmetic facts (Campbell & Graham, 1985, Campbell, 1987b). It is also the case that not all information is retrieved with the same level of ease; the problem-size effect, for instance, suggests that some parts of the network (involving smaller-numbered operations) are more accessible than other parts of the network (Campbell & Graham, 1985).

There is also evidence for the use of strategies other than retrieval from a structured network to solve arithmetic problems (LeFevre et al., 1996), especially in developmental contexts (LeFevre, Smith-Chant, Hiscock, Daley, & Morris, 2003). Early models of decision making in verification tasks suggested that the processes involved took the form of "retrieve-compare", wherein an initial obligatory stage of answer retrieval was followed by answer comparison and response execution stages (Ashcraft & Battaglia, 1978). However, current behavioral evidence suggests that there are multiple routes to judging answer correctness in verification tasks, including the use of familiarity-based judgments (Zbradoff & Logan, 1990; Lochy, Seron, Delazer, & Butterworth, 2000; Campbell & Tarling, 1996), the strategic use of rule-based information like odd/even parity matching (Lemaire & Reder, 1999), as well as the use of direct comparison to retrieved answers (Ashcraft, Fierman, & Bartolotta, 1984).

In general, the accumulated evidence suggests that behavioral results obtained in these verification tasks (i.e., delayed RTs) are the result of a sum of many internal cognitive processes, which leads to inferential difficulties. For example, are the reaction time delays due to conflicts at the level of responses or decisions, or are these memory confusions operating during even earlier stage of initial semantic long-term memory access?

Investigations of the stages of processing involved in accessing long-term memories for arithmetic facts has been inspired by how language researchers used ERPs to approach similar questions in the domain of sentence processing, where there are also different kinds of comprehension difficulties as a result of the presentation of unexpected sentence-final words. A series of ERP findings will be reviewed, starting with an overview of the basic sentence processing literature and moving towards applications most relevant for addressing the question of why certain answers to arithmetic problems are harder to correctly reject as incorrect.

N400 effects in sentence comprehension

It is the case that, as coherent full sentences unfold, the N400 amplitude at each successive word linearly decreases on average. This implies that the amount of new semantic information that is revealed by each word is somewhat smaller than that of the previous word, so long as a sentence is unfolding such that it conveys a reasonable message (Payne, Lee & Federmeier, 2015; Van Petten & Kutas, 1990). By the final word of a coherent sentence, then, expected endings have very facilitated (small) N400 responses.

When an unexpected word is encountered, it demands relatively more new semantic information to be accessed, and the N400 response is larger as a result. Although it is tempting to view this large N400 response to unexpected words as unusual, it is actually similar in magnitude to what might be seen in response to out-of-context single words that were provided no support, like words at the beginning of a sentence (Kutas & Federmeier, 2011).

Beyond the extremes of expected versus unexpected sentence completions, the N400 amplitude is more generally inversely related to the probability that participants would choose to end a sentence with the given item (a measure known as cloze probability). For example, given the sentential context of, “He caught the pass and scored another touchdown. There was nothing he enjoyed more than a good game of _____”, the word “football” has a higher cloze probability than does the word “monopoly”. The N400 is consequently smaller (more facilitated) in response to the expected word “football” compared to the unexpected word “monopoly” (Federmeier & Kutas, 1999b). The most common account of this phenomenon is that it is due to the brain being more prepared to process the more congruous, expected input (Kutas & Federmeier, 2011).

Even when words are matched for cloze probability, however, other factors can influence how prepared the system is to process them. For example, “baseball” is just as unexpected an ending as “monopoly” was in that example sentence (in norming studies, few to no people choose to end the sentence either way), but the N400 response to “baseball” is nonetheless reduced or facilitated compared to “monopoly”. The common explanation of this “related anomaly effect” is that this facilitation is the result of semantic overlap between the words “football” and “baseball” – they are related to one another in long-term semantic memory due to shared category membership (e.g., both are team sports involving athletics). Thus, even though “baseball” is technically unsupported by the context, it does not require as much *new* semantic information to become active as an unexpected item from a different, semantically unrelated category like board games does (Federmeier & Kutas 1999a; see also Thornhill & Van Petten, 2012 for another case of N400 facilitation due to semantic relatedness).

Late positive complexes in language comprehension and memory

Following the N400, there are also often positive-going waveforms (generally referred to as making up the Late Positive Complex, LPC) that are typically larger in cases where some kind of additional cognitive processing (e.g., more memory retrieval) occurs. At the word-level, a late positivity emerges after presentation of words that are more difficult to pronounce (e.g., "yacht") relative to words that are more transparent in their spelling-to-sound correspondence (e.g., "saw") when subjects are required to read the items out loud (Fischer-Baum et al., 2014). There is also an LPC in word repetition paradigms, with a larger LPC in response to repeated low-frequency words compared to repeated high-frequency words (Rugg, 1990). During recognition paradigms in which subjects make old/new judgments on (potentially) previously seen items, those items correctly identified as old (often) elicit larger LPCs than do new items (e.g., Rugg & Nagy, 1989; Smith & Halgren, 1986; Van Petten & Senkfor, 1996).

At the sentence level, there are larger LPCs in response to some kinds of unexpected endings and not others. In 2012, Van Petten and Luka reported that LPCs were present (if not reported) in approximately 33% of studies using sentence stimuli. Given the variety of eliciting conditions across the studies, how the LPCs were interpreted varies substantially at a mechanistic level, but it is generally agreed that positivities emerge when a stimulus is encountered that requires more extensive processing – for example, the retrieval of relatively more information from memory. There are also views that these later positivities are just a manifestation of the P300b, a component that emerges for task-relevant, probabilistically unexpected events and is linked to the updating of context (Donchin & Coles, 1988; Coulson, King & Kutas, 1998; Coulson, 1998).

N400 effects across modalities, and earlier negativities

Given that these effects are interpreted at the level of semantics, which is defined fairly broadly to encompass any number of representational features, a reasonable question is how generalizable the findings are to other modalities. The answer seems to be that they generalize well: sentences presented auditorily, and sentences that are completed with line drawings instead of words, replicate the general pattern with most facilitation for expected items, intermediate facilitation for unexpected items from the same category as the expected items, and least facilitation for unexpected items that are not from a different category (Federmeier & Kutas, 2001; Federmeier, McLennan, de Ochoa & Kutas, 2002). However, these N400 effects often overlap or co-occur with other negativities that are more modality-specific in their functional properties. Even in word recognition studies there is a prior component, the N250, which is linked to processing at the orthographic, rather than semantic, level of analysis (e.g., Dickson & Federmeier, 2014). Similarly, in sentential completions and repetition designs using line drawings or pictures, the N400 congruity effect is preceded by earlier effects on a negative-going waveform that is often called the N300 (e.g., Federmeier & Kutas, 2001).

Characterizations of the response properties of the N300 have reported that it patterns much like the N400, showing facilitation upon direct repetition or following semantic priming (Barrett & Rugg, 1990; McPherson & Holcomb, 1999). However, it has been more linked with the processing of higher-level object information than to the actual semantic content of the object per se (Schendan & Kutas, 2003; Schendan & Lucia, 2010).

Additionally, although in sentential contexts it has been found to respond in a graded manner, with facilitation for semantically related picture endings compared to unrelated endings (Kutas & Federmeier, 2001), in designs wherein pictures are not presented following sentential contexts, the N300 has not been found not to be as sensitive to this level of semantics (e.g., in repetition priming, Holcomb & McPherson, 1999). The specifics of the negativities preceding the N400 are thus not as well defined as the N400 itself is, mostly because they appear under more restricted contexts (e.g., during picture viewing, face viewing, or during presentations of stimuli to either visual field rather than in full central vision), but generally they are thought to be part of the overall attempt to take an input and connect it to knowledge of perceptual regularities (Federmeier, Kutas & Dickson, 2015 for a review of negativities preceding the N400 across modalities).

Contextual effects in sentences and equations: direct comparison

Researchers interested in the domain of numerical cognition have used the findings across modalities in the language and object recognition literatures as a model for how to use ERPs to answer their own questions about long-term memory access during arithmetic problem solving. Rather than having sentences set up a context, they instead have used arithmetic problems and examined the brain response across types of suggested answers (i.e., comparing correct, incorrect, and confusing answers instead of unexpected and expected sentence-final words).

The question of how the effect of sentential congruity compares directly with the effect of answer correctness was addressed by one of the first studies to examine the ERP response to answers in the context of arithmetic problems (Niedeggen, Rösler, & Jost, 1999). In a within-subjects design, the effect of congruity in sentences (responses to contextually expected/unexpected words) and the effect of correctness in numerical expressions (correct /incorrect answers to multiplication problems) were directly compared. They reported that both sentences and equations have a near-identical response pattern (N400 effects) when comparing effects of congruity in final words/answers.

Studies of this kind have also looked at the response to incorrect answers that are known to be behaviorally confusing, which are conceptually similar to unexpected words that are category-related (e.g., "baseball" when expecting "football"). As previously described, these table-related answers are derived by altering one of the existing operands in a multiplication problem. Given the operation "6 x 7", the answer "36" is harder to categorize as being incorrect than is "40", because "36" is the correct answer for the very similar operation of "6 x 6". In ERPs, the related/incorrect answers elicited intermediate effects between correct and incorrect answers, including facilitation on the N400 compared to incorrect answers that were not related to the correct answer, so long as the incorrect/related answers were not too numerically distant from the correct answer (Niedeggen & Rösler, 1999; Niedeggen, et al., 1999). These results were interpreted as an indication that the semantic activation of correct answers spreads to similar/related answers at relatively early stages of processing (i.e., during initial contact with long-term memory), and that the source of associative confusions during production and verification is not only at downstream stages of responding and decision making.

It should be noted, however, that this literature is currently undergoing some skeptical revisions. For example, Jasinski and Coch (2012) measured ERP responses to correct/incorrect solutions across all four major arithmetic operations. They reported a peak around 280ms (identified as belonging to the family of pre-N400 negativities like the N300) that was sensitive to answer expectancy (correctness), a P300b specific to correct answers (likened to target detection), and an LPC that was largest for incorrect answers and that they connected to plausibility assessment and double-checking. Notably, they did not find an N400 effect of correctness/expectancy. The authors argue that the previously reported N400-like effects might have been an artifact of using difference waves (which are generated by subtracting correct equation completions from incorrect ones) to identify the component (e.g., by Niedeggen, Rösler, & Jost 1999). Specifically, their claim is that the "N400"-looking effect was artificially created through a mix of the P300 (larger positivity for correct answers) and N280 (larger negativity for incorrect answers) effects across conditions of correctness. Visual inspection of the original reports is consistent with this interpretation, although the plots are limited to few channels, and it is difficult to assess without full access to the original data.

So what would this mean at a cognitive level? If the amplitude of the N400 is typically thought to reflect new semantic information becoming active, and if there is no difference on this component in response to different types of arithmetic answers, then it would seem to suggest that mathematical knowledge may be different from other kinds of general semantic information that gets accessed when meaningful items are being processed. The N280 might be more similar to the family of negativities seen prior to the N400 in picture-processing tasks like the N300, which also show effects of expectancy and congruity. Given that N300 effects are almost always directly followed by N400 effects, and that they often pattern in very similar ways, it is unclear what they would mean in isolation, especially given that their previous eliciting conditions have been in response to line drawings rather than symbolic numbers.

More work is thus necessary to definitively characterize the waveforms of interest in order to better understand the cognitive mechanisms involved in arithmetic. This forms the motivation for Experiments 1 and 2. In Experiment 1, I will first attempt to replicate the original report of relatedness effects in arithmetic on the N400 component using modified stimuli and a more robust sample (the original reports used 12-16 participants). Experiments 1 and 2 will further examine these effects by exploiting differences in the neurobiological organization of brain structures involved in arithmetic processing versus language processing. Specifically, whereas there are well-known differences in how *language* processing unfolds across the cerebral hemispheres, the potential for hemispheric differences is not typically considered in *arithmetic* processing.

The evidence for these neurobiological differences in arithmetic and language processing, and how they can be applied to better distinguish between the ERP effects of interest, are reviewed below.

Brain network in numerical cognition (arithmetic)

In an early, foundational study of the areas of the brain involved in number processing, four critical tasks were compared: multiplication, subtraction, digit naming, and number (size) comparison. All four tasks commonly activated a bilateral fronto-parietal network (relative to a baseline letter naming control task), with the frontal part being attributed to working memory and executive control processes (not specific to arithmetic) and the parietal involvement more specifically attributed to number/mathematical processing (Chochon, Cohen, Moortele, & Dehaene, 1999).

This fronto-parietal network engaged by arithmetic and number processing has been replicated many times, although more careful distinctions are now made between the intraparietal sulcus (IPS) activity and the activity in the superior parietal lobe (SPL) (see, e.g., Rosenberg-Lee, Chang, Young, Wu & Menon, 2011). One group applied TSM to the IPS and SPL of either cerebral hemisphere while participants solved arithmetic (multiplication and subtraction) problems, and they found both longer RTs and more errors as a result of TMS to either right or left IPS (but no effect of TMS to the SPL). Notably, the higher errors in multiplication were due to subjects providing table-related incorrect answers (Andres, Pelgrims, Michaux, Olivier & Pesenti, 2011). Since IPS activity being disrupted resulted in poorer outcomes for both multiplication and subtraction, the processing performed in this region – in both hemispheres – was interpreted as serving a more functionally causal role.

This largely bilateral fronto-parietal network for arithmetic stands in contrast with written language processing in two fundamental ways: first, the language-processing network is traditionally more left-lateralized, whereas the findings with arithmetic in the parietal lobe suggest more bilateral involvement, and secondly, language processing is generally associated with a fronto-temporal network (with the parietal lobe playing different roles in different accounts, e.g., Lau, Phillips, Poeppel, 2008; Price, 2012).

Relatively little deliberate attention has been paid to hemispheric differences in arithmetic ability, but the many reports that do exist suggest there are substantial differences (cf., Andres et al, 2011). The precise nature of the capabilities of each hemisphere remains elusive because of differences in skills being tested (e.g., arithmetic knowledge, magnitude judgment abilities, processing of Arabic digits, dot pattern assessments) and because of differences in the methods used to derive inferences: lesions (e.g., Warrington, 1982), callosotomized patients (e.g., Funnel, Colvin, Gazzaniga, 2007), and visual field manipulations in non-brain damaged individuals (e.g., Katz, 1980; Klein & McInnes, 1988; Castro, Sumich, Premkumar, & Jones, 2014).

These findings, which uniformly suggest that differences exist (although they disagree about what the differences *are*), have not permeated the broader field of number cognition the way that processing differences across the hemispheres in language have. Although there has been no use of ERPs to study hemispheric differences in arithmetic processing, a review of language research in this domain will serve to demonstrate the utility of the approach of using ERPs to study hemispheric differences during long-term memory access.

Hemispheric differences in sentence processing: evidence from ERPs

The known biases in language processing to the left hemisphere of the brain has motivated interest in examining hemispheric differences in sentence processing, and many of the previously discussed EPR effects at the level of both words and sentences have been found to be affected by manipulations that bias the hemisphere responsible for processing the input (e.g., at the sentence level, Federmeier & Kutas, 1999b, 2002; at the word level, Dickson & Federmeier, 2014). This work employs the method of visual half-field presentation, which biases initial processing of an input either to the left hemisphere (stimulus presented to the right visual field, LH/RVF) or to right hemisphere (stimulus presented to the left visual field, RH/LVF). For a review of the efficacy of this technique and comparison to "split-brain" patient work, see Banich (2002). Interhemispheric transfer of some kind is likely. However, inferences typically rely on the presence/absence of effects across hemispheres, which are difficult to explain if one hemisphere is really communicating all of its information to the other. The information that does transfer from the stimulated visual field/hemisphere appears to be substantially impoverished.

In order to compare how each hemisphere uses context to activate semantics for upcoming words, participants read the sentences that have been previously described in central presentation as normal (e.g., "He caught the pass and scored another touchdown. There was nothing he enjoyed more than a good game of _____"), but then the ending types (e.g., expected: "football", unexpected but within-category/related: "baseball", and unexpected/unrelated: "monopoly") were lateralized to one visual field or the other, seemingly at random. Interestingly, the intermediate N400 facilitation for the unexpected/category-related items was only present in the LH/RVF condition. The RH/LVF showed no difference between the different kinds of unexpected items. This was interpreted as suggesting that the RH does not activate the semantic network of the expected word prior to it being presented (Federmeier & Kutas, 1999b).

When the lateralized completions to the same sentences were in the format of pictures, the same pattern of results across the hemispheres obtained: whereas the RH did not distinguish between types of unexpected items (related/unrelated), the LH did show N400 facilitation for category-related unexpected endings (Federmeier & Kutas, 2002). The N300 component did not show this effect of lateralization, instead showing facilitation for the category-related unexpected endings in both visual fields. The general conclusions from this line of work are that the hemispheres access semantics through different mechanisms during online sentence comprehension (for review, see Federmeier, 2007).

The specific question being addressed in Experiment 1, then, is what happens when different types of *arithmetic* answers are presented in either VF. Do the findings of LH-specific facilitation for related items replicate, and, if so, on what particular ERP component (N300? N400? LPC?)? Given the anticipated results (i.e., that there will be some kind of difference across the hemispheres), Experiment 2 – a purely behavioral experiment – is then necessary to interpret the practical significance of the findings.

Symbolic number processing out of arithmetic contexts

So far, the proposed work seems to assume that the most interesting and relevant way to approach the study of the long-term memory representation of symbolic numbers is in the context of arithmetic processing. In daily life, this is probably not actually when adults most often encounter and use symbolic numbers. They are also used during more generic measuring/counting tasks (e.g., “2” spoons of sugar, denoting speed limits) or as unique identifiers (e.g., credit card numbers, confirmation numbers), among other contexts (e.g., sports jerseys) that do not directly involve arithmetic calculations. When people view symbolic numbers in these other contexts, what is the nature of the representation being activated?

It is clear from my prior work with word stimuli that the goal or emphasis of the task influences what properties of stimuli are brought to mind – i.e., only when participants actually had to name a word out loud were effects of spelling-to-sound regularity observed (in the form of an LPC to low-frequency irregular words like "yatch"). When participants were just reading the list of items to detect proper names, no LPC to these same items was observed (Fischer-Baum et al., 2014). Also, as previously reviewed, depth of encoding influences the size of repetition effects (e.g., Rugg et al., 1988). One particular finding was that items that typically fail to elicit repetition effects (non-meaningful strings of unpronounceable letters like "TXRM"), *do* elicit repetition effects when the task forces them to be assessed more carefully – in the case of Laszlo, Stites and Federmeier (2012), this was found by having participants perform a modified lexical decision task in which meaningful acronyms (which are also strings of unpronounceable letters) were a potential target.

Taken together, one implication of these results is that symbolic representations of number like Arabic numerals should be able to elicit a repetition effect, when they are viewed in a list format one-by-one, although no studies to my knowledge have deliberately attempted to demonstrate this systemically. It is also possible that the overall magnitude of the effect might differ as a function of task. Notably, no one has tried to characterize the general properties of ERPs to these numerals outside the context of arithmetic problems. We already know that, under some contexts, there is an N400 effect of frequency in the ERP response to words. Of interest, then, is whether or not there a similar effect (on the N300, N400, or LPC) in response to numerals, instead sensitive to the amount of quality of experience an individual has with a given numeral. These questions motivate Experiments 3 and 4, which will further explore how symbolic numbers are represented in long-term memory, but using a repetition design with multiple tasks instead of embedding numbers only in the context of arithmetic expressions.

Summary and current studies

In the event-related potential (ERP) literature on the processing of mathematical expressions, there is currently controversy over whether or not, or to what extent, their completions (the answers) are processed similarly to how final words are processed in sentences – with specific interest in whether effects of contextual congruity are on the N400 component or not. Additionally, in the broader literature of cognitive neuroscience there are reasons to suspect that mathematical skills and knowledge are distributed differently across the hemispheres of the brain than are the skills and knowledge relevant for language use. In particular, whereas language is generally dominated by left-lateralized regions in a fronto-temporal network, mathematics (and the more general "number sense") seems to be more distributed across *both* hemispheres in a fronto-parietal network.

In Experiment 1, therefore, the present work will test the boundaries of just how similar the processing of numbers in context is compared to words in sentences by biasing processing of numbers across the hemispheres (via visual half-field presentation) to see whether the patterns in each visual field/hemisphere that have previously been reported for word stimuli also appear for correct (expected) answers, incorrect (unexpected/unrelated) answers, and incorrect table-related (unexpected/related) numerical stimuli.

At a behavioral level, although there is substantial research on how answers to arithmetic problems are retrieved, there is seemingly no accompanying literature in healthy populations on hemispheric differences in these particular skills. In language research, it is generally found that the left hemisphere “wins” when presented with verbal challenges of many kinds (e.g., Babkoff & Faust, 1988; Jordan, Patching, & Milner, 2000). In order to draw inferences about the results from the ERP findings, a basic behavioral measurement of skill assessment in the same task is useful. Thus, in Experiment 2, participants will be presented with the same materials as in Experiment 1, but will instead have their reaction times and accuracies measured.

When assessing the ERP response to Arabic numerals in context, there has rarely been joint consideration of the properties of the out-of-context numerals as potentially meaningful objects by themselves. This is in contrast with language research, which has a long history of focusing on out-of-context individual words (which nevertheless always have a broader *task* context). From that work, the field knows that a number of properties influence the basic ERP response to words themselves, such as frequency and orthographic structure, and this knowledge allows researchers to make inferences about how word knowledge is organized in the brain. Similarly, effects of priming and repetition at the word level provide indications for how memories for these words are accessed and related to one another. Knowledge about how information associated with numerals is structured and accessed from memory is comparatively limited, and, in order to understand how such information is accessed in the context of expressions (Exp. 1-2), it is necessary to better understand the nature of how meaning processing unfolds for the out-of-context representations.

Therefore, in Experiment 3, numerals will be presented in a repetition design in order to characterize the basic properties of responses to them. In word recognition studies, it is already known that both initial responses and repetition patterns differ across types of task. Participants' attention will be therefore be focused at different levels of analysis of Arabic numerals: in a task focusing on visual properties (the "matching" task), in a task emphasizing arithmetic fact retrieval (the "divisor" task), and in a task requiring conceptual magnitude assessment (the "quantifier" task). The results of Experiment 3 will in turn strengthen the interpretability of Experiment 1, in part because it will allow an examination of the ERP response to numerals *without* the presence of decision-related components (e.g., P300), as the critical task manipulations will always be on stimuli that are only seen after the initial, target number presentation.

Finally, Experiment 4 will address the possibility of non-uniformity in numeral processing at an item level. The same basic design as Experiment 3 will be used (numerals incidentally repeating), except that participants will perform a familiarity judgment task (i.e., a "yes" judgment might be given because it is their apartment number). This allows for repetition effects on the N400 and other memory retrieval components (i.e., LPC) to be compared within a single task but across item types whose long-term memory representations might vary.

CHAPTER 2

Experiment 1: Hemispheric differences in arithmetic fact evaluation as revealed by ERPs

Introduction

The most visible characteristic of the organizational structure of the human brain is that it is made up of two cerebral hemispheres. Although their cellular and broader neurobiological makeup is very similar, cognitive functions are often strongly lateralized to one hemisphere or the other, and this division of labor between the hemispheres seems to be an important principle of typical neural functionality. The first report of such cognitive lateralization –for language production in the left hemisphere – was integral to the foundation of neuropsychology as a field (Broca, 1865). Interest in lateralization of cognitive specialties has since expanded outside of the language domain, with research spanning broad topics such as hemispheric differences in sensitivity to different spatial frequencies (Sergent & Hellige, 1986; Christman, Kitterle, & Hellige, 1991), in attentional biases in global versus local features of an object (Martin, 1979), and in bottom-up versus top-down processing (Federmeier, 2007), among others. Here, our report will focus on another domain of research that has been devoting increasing attention to the processing mechanisms and abilities of each hemisphere: numerical cognition, and arithmetic processing in particular.

An interesting analogy can be formed between sentence reading (language) and arithmetic expression reading (math), which might suggest shared underlying cognitive processes. Sentences are made up of subparts (words) that are systemically combined to convey a potentially coherent message, and, similarly, arithmetic equations are made up of subparts (numbers) with combinatorial symbols (+, %, x, -) that can be sensibly completed (e.g., 4 x 5 =

20) or not. At the same time, there is clear evidence that there are important differences in the neural systems that are engaged in these two processes: at a whole-brain level, whereas sentence and word reading is largely associated with fronto-temporal activity (Lau, Phillips, Poeppel, 2008; Price, 2012), mathematical expression and number processing is more associated with fronto-parietal activity (Chochon, Cohen, Moortele, & Dehaene, 1999). Although the specific subregions involved seem to differ across math and language, it remains unclear to what extent the two share similar patterns of hemispheric lateralization overall, given the complexities of the subprocesses involved and the somewhat mixed evidence in the existing literature (described below). Therefore, the aim of this study is to use a combination of techniques novel to this field to examine the lateralization of function for arithmetic and to then compare it with prior results using the same techniques to study language lateralization.

Within the domain of mathematical cognition, the dominant perspective is that lateralization of function and its relationship to language depends heavily on the numerical skill being tested (Dehaene & Cohen, 1995). If the skill is retrieval of exact facts from memory (e.g., during multiplication), then this has been linked to verbal abilities, and has been more associated with language processing and the left hemisphere. However, if the numerical skill being tapped is instead more about approximation or basic number comparison (e.g., judging the size of relative quantities), then there is less association with language, and, in turn, more association with bilateral function. The evidence for each of these claims follows, starting with the relationship between lateralization of language function and memorized arithmetic facts (typically multiplication), and then proceeding through the evidence that more general numerosity concepts can be independent of left-lateralized language abilities.

In the particular domain of mathematical expression processing, calculation impairments in patients with unilateral brain lesions have been relatively more associated with damage to the left hemisphere (Jackson & Warrington, 1986, Rosselli & Ardila, 1989; Ashcraft, Yamashita, & Aram, 1992). Furthermore, in patients with severed corpus collosums (i.e., “split-brain” patients), when the right hemisphere alone is forced to perform arithmetic, severe impairments are reported, particularly in multiplication (Gazzaniga & Smylie, 1984; Funnell, Colvin, & Gazzaniga, 2007). In contrast, the isolated left hemisphere is typically able to perform these arithmetic tasks above chance. When patients with left hemisphere lesions were specifically examined for extent and type impairments to their language and numerical skills, it was found that more severe language impairments were correlated with impairments in quantitative abilities – again, particularly for multiplication (Delazer, Girelli, Semenza, & Denez, 1999). This evidence forms the basis for the argument that arithmetic fact retrieval is dependent on left-lateralized language processing abilities.

That language processing, and production in particular, tends to be lateralized to the left hemisphere is one of the most well-known features of brain organization (e.g., Geschwind Levitsky, 1968). However, in many individuals, language production is instead lateralized to the right hemisphere (Binder et al., 1996; Bishop, 2013). For neurosurgeons, identifying the lateralization of language in a patient often forms a critical part of their presurgical assessment process, which is achieved through a procedure known as the Wada test (in which amobarbital is injected into each of the left and right carotid arteries in turn, to examine the abilities of the unaffected hemisphere, Wada & Rasmussen, 1960). An interesting question, then, is what happens to arithmetic skills when language is lateralized to the right hemisphere?

When patients for whom the lateralization of both language and arithmetic skills were both assessed, the lateralization of language influenced the arithmetic abilities of each hemisphere. Specifically, if language was left-lateralized, then the isolated right hemisphere was impaired at performing multiplication, whereas, if language was right-lateralized or bilateral, then the isolated right hemisphere was above chance at multiplication (Delazer, Karner, Unterberger, Wasler, Waldenberger, Trinka & Benke, 2005). This suggests that it is not just a coincidence that both language and arithmetic skills tend to be left-lateralized. Instead, they seem to track each other, either due to a shared reliance on a higher-level processing mechanism that itself tends to be lateralized, or because arithmetic skills actually depend on language abilities (as is suggested by the association between level of language impairment and level of arithmetic impairment).

The most extreme versions of this conclusion – that, without language, there can be no arithmetic fact knowledge, and that the contralateral hemisphere has no involvement in arithmetic processing – is unlikely and must still contend with evidence for dissociations between these abilities. For example, there is evidence from a TMS study that both hemispheres are causally involved in answer generation for multiplication problems (Andres, Pelgrims, Michaux, Olivier & Pesenti, 2011). There has also been a long history of reporting case studies of patients for whom language and arithmetic abilities are dissociable. In several cases of patients with semantic dementia, which is a progressive neurodegenerative disorder featuring loss of semantic memory (especially in word comprehension), there are reports of successful retention of some (or all) arithmetic skills (Diesfeldt, 1993; Cappelletti, Butterworth, & Kopelman, 2001, 2012; Crutch & Warrington, 2002). However, even in these reports, it is often the case that multiplication is the most impaired arithmetic ability when verbal memory is compromised. Thus, although there is some evidence implicating right hemisphere involvement in production of multiplication problem answers, there seems to be a strong relationship between left hemisphere language abilities and the production of answers to multiplication problems.

Outside of the domain of producing exact answers to arithmetic problems from verbal memory, and in the domain of symbolic and non-symbolic number comparison, the dependence on left-lateralized language abilities is much less apparent. Instead, there are reports that both hemispheres can compare symbolic numbers for their relative size (Colvin, Funnell, & Gazzaniga, 2005). Indeed, there is evidence directly linking developmental changes in subregions of the right hemisphere's parietal lobe with successful acquisition of symbolic and non-symbolic magnitude judgment abilities (Holloway & Ansari, 2010), and evidence from TMS demonstrating that disruption of these right hemisphere areas results in lower performance on tests of automatic magnitude judgment abilities (Cohen Kadosh et al., 2007).

Along with magnitude judgments, the right hemisphere also seems to be able to engage in generating answers to arithmetic problems through approximation or through step-by-step deliberative procedures. For example, the same left hemisphere lesion patients who can fail to report multiplication answers often retain explicit knowledge of arithmetic operations and the ability to deliberately apply this knowledge in addition and subtraction (e.g., Cohen, Dehaene, Chochon, Lehéricy, & Naccache, 2000). In one case study, errors in addition and subtraction were only small distances from the correct answer, suggesting that they were produced through tracking and manipulating conceptual quantities and magnitudes (Funnell, Colvin, & Gazzaniga, 2007).

In sum, it would appear that the right hemisphere's numerical ability is limited to answer approximation and magnitude judgments, leaving the performance of exact recall of mathematical arithmetic facts as the domain of the left hemisphere (Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). However, there are still several aspects of this research field which remain relatively unexamined: (1) the specializations of the hemispheres seem to rely on skills that unfold under different timescales, but no time-sensitive neural measures are reported, and (2) the reports at the level of the hemisphere (i.e., in commissurotomized patients) rely on some kind of end-state explicit report from the participant, which renders failures to succeed difficult to interpret because it is unclear at what level of processing the failure occurred. In general, these gaps leave it ambiguous whether, for example, the right hemisphere generates exact correct answers and then undergoes another process that renders those answers difficult to access and report, or whether the right hemisphere simply is not a reliable independent source from which to generate the exact answers to multiplication problems.

At present, since there are no studies using time-sensitive measures to understand how the processing of equations dynamically unfolds across the hemispheres, these questions remain open. Addressing this gap is critical given how quickly the answer evaluation process occurs – people are able to verify the correctness of simple arithmetic problems easily after less than a second of exposure to the problem, but whether there are differences in exactly when and how this occurs in each hemisphere is currently undetermined. Fortunately, a successful and productive technique for addressing this type of question has already been developed in order to answer similar questions in the language domain.

Indeed, over a decade of research has combined visual half-field presentation with event-related potentials (ERPs, a time-sensitive neural measure) to examine how the hemispheres differently apprehend words for meaning under different types of priming conditions (Kandhadai & Federmeier, 2010; Dickson & Federmeier, 2014) and different types of sentential contexts

(Federmeier & Kutas, 1999b; Wlotko & Federmeier, 2007). In the visual half-field technique, items of interest are presented to subjects lateralized to either visual field, which provides processing advantages to the contralateral hemisphere (i.e., items presented in the periphery of the left visual field will be preferentially apprehended by the right hemisphere, and vice verse for items in the right visual field). Although the hemispheres are still connected via the corpus collosum, such that interhemispheric communication will eventually be possible, the advantage afforded by direct stimulation is substantial and what information does become shared is both delayed and degraded (see Banich, 2002, for review).

Under central presentation conditions, a handful of studies have used ERPs to examine arithmetic processing as it unfolds, typically with an equation presented item by item, followed by an answer that is presented for verification (e.g., Niedeggen & Rösler, 1999, Niedeggen, Rösler & Jost, 1999; Jasinski & Coch, 2012). Researchers have then used this paradigm to examine the timecourse with which the provided answer's fit to the prior equation is assessed, enabling them to answer questions about when and how the system appreciates that the answer was correct or incorrect. The present study's approach will be to use this same procedure while also lateralizing the answers to either visual field in order to uncover possible differences across time in the hemispheres' sensitivity to answer types in arithmetic problems.

At present, there are no such studies using this approach to examine how the timecourse of mathematical answer evaluation proceeds within each of the two hemispheres. The current paper's primary goal, then, is to begin to fill this gap in the literature by recording ERP responses to answers to multiplication problems when the processing is biased to either hemisphere. Multiplication was chosen as the operation since we wanted to draw comparisons with hemispheric differences in language comprehension, and this operation is most closely associated with verbal language processing (as previously reviewed). Due to the relative novelty of the particular manipulations and stimulus set, we will also report the ERP responses to the same answers in standard, central visual presentation as well.

Since no one has examined responses to any arithmetic problems with ERPs in either visual field before, we will describe what has been found in language studies in each hemisphere to provide a basis of comparison. In ERP studies of language processing, each of the hemispheres have consistently demonstrated that they can distinguish between expected and unexpected sentence-final words (Federmeier & Kutas, 1999a, Federmeier, Mai, & Kutas, 2005). If the verbal retrieval that occurs during multiplication is similar, then this suggests that both hemispheres might similarly be able to categorize correct from incorrect answers to arithmetic problems, counter to the current suggestion that only the left hemisphere has access to these exact answers.

However, the effect of correctness might not manifest on the same component in arithmetic and sentence processing. In language studies, this effect is seen on a component linked to semantic memory access, the N400. It is unclear whether clear N400 effects obtain in arithmetic processing: although early work identified such effects as an N400, more recent studies have made a strong case that the effect is actually driven by a form of the P300, which instead reflects categorization and updating of working memory and is larger (more positive) for target items (Donchin & Coles, 1988; also see Jasinski & Coch, 2012; for an elegant re-examination of the component identification in this paradigm; and, for an example of a typical-looking target P300 response to correct answers when participants have sufficient time to generate the answer, see Fig. 2 in Domahs, Domahs, Schlesewsky, Ratinckx, Verguts, Willmes, Nuerk, 2007, and compare it to Fig. 1, when they participants have less time to prepare for the answer). Even in cases where sensitivity on a negative-going component that resembles the N400 appears to be the case, there is a following positivity that renders analysis of the prior negativity ambiguous (see Fig. 5 of Niedeggen, Rösler, & Jost, 1999, where only one scalp channel is plotted).

Following this interpretation, we expect to see a P300 effect related to correctness and will be focusing on its amplitude and timing (which is linked to stimulus evaluation time and, as such, tends to correlate with reaction times) (Kutas, McCarthy & Donchin, 1977). A relevant first question to address, then, is whether both hemispheres categorize correct answers as correct (and distinct from incorrect answers that are more easily confusable with correct answers), and, if they do, whether this categorization completed at the same speed in each hemisphere, as indexed by P300 latency.

Whereas basic appreciation of congruity seems to be similar across the hemispheres in language (and remains to be examined in the context of arithmetic), other manipulations on sentence-final completions have revealed striking hemispheric differences. In particular, subtypes of unexpected sentence final words are processed differently in the two hemispheres. When unexpected endings that came from the same semantic category as the expected ending (e.g., "apples" when the expected ending was "oranges") were presented as well as unexpected endings that came from a different semantic category as the expected ending (e.g., "carrots", which are not as closely related to "oranges"), only under right visual field/left hemisphere (RVF/LH) presentation conditions was there facilitation on the N400 for the semantically related unexpected ending. In the left visual field/right hemisphere (LVF/RH), the unexpected, within-category item ("apples") was processed similarly to the across-category item ("carrots"), consistent with an interpretation in which only the left hemisphere pre-activated the semantics including (and surrounding) the most expected final word. In sum, to the LVF/RH, both subtypes of unexpected items were distinguished from the expected items (and were not different from each other), whereas the RVF/LH distinguished between all three types of completions.

In the interest of comparing these kinds of verbal memory processes across domains, it was of interest to include two types of unexpected (incorrect) answers, one which might be more related in long-term memory to the expected ending than the other. Drawing on the existing multiplication fact retrieval and production literature, which suggests that knowledge of multiplication answers is organized in a table-like fashion (wherein multiples of a given operand relatively more confusable with each other than with multiples of another operand), we included both table-related and table-unrelated incorrect answers, to be compared with correct answers. With this manipulation, incorrect related answers are correct answers for a different problem that shares an operand with the presented problem (e.g., given the problem setup “6 x 7”, an incorrect related answer would be “36”, the correct answer to “6 x 6”), whereas incorrect unrelated answers do have have property (e.g., given the problem setup “6 x 7”, an incorrect unrelated answer would be “40”, which is not evenly divisible by 6 or 7 but is still numerically close to the correct answer of “42”). In production studies, the most common type of error is this type of incorrect related answer (which also take longer to reject than the incorrect unrelated answers in verification paradigms), presumably because of their heightened confusability in long-term memory with correct answers (Campbell & Graham, 1985; Campbell, 1987).

Neural evidence for how sensitive each hemisphere is to this type of manipulation in multiplication is limited, and in particular there is no literature (to our knowledge) on how patients who have undergone commissurotomies respond in their separate hemispheres to these answer types. If one takes the existing evidence to indicate that *all* multiplication fact knowledge is relatively more left-lateralized, including table-relatedness, one might expect that a similar manipulation of subtypes of incorrect (unexpected) answers would find only RVF/LH differentiation between incorrect answers that are unrelated to the context generating the correct answer and incorrect answers that are highly related to the context generating the expected correct answer.

Additionally, at a mechanistic level, the left hemisphere relatedness effect in language is often argued to reflect the pre-activation of the expected ending. If the same predictive mechanism is at play during arithmetic processing, then this reasoning would also predict a relatedness effect for arithmetic answers selective to the left hemisphere. On the other hand, some of what the right hemisphere has been argued to do in arithmetic tasks is more methodical and deliberative – breaking down the problems into steps since direct retrieval mechanisms seem not as effective (e.g., Funnell, Colvin, & Gazzaniga, 2007). This kind of processing could potentially also lead to a relatedness effect, but it would likely unfold more slowly across time. Thus, identifying when and how the hemispheres might differently distinguish between subtypes of incorrect answers forms not only an important inroad to relating lateralization during arithmetic processing with lateralization during language processing, but also will provide novel information about the nature of relatedness effects and contextual congruency in arithmetic processing in each hemisphere in general.

Once again, the precise component that might show this differential effect across hemispheres is again unlikely to be the same in arithmetic and language. When incorrect unrelated answers have been presented in ERP studies in central vision, the majority of the effect manifests after the initial P300, in the form of a late positivity for incorrect unrelated items relative to incorrect related item. Again, initial reports of this identified the effect as arising on the N400 and drew inferences about the structure of semantic memory for table-relatedness (Niedeggen & Rösler, 1999; Niedeggen & Rösler, & Jost, 1999), however, the co-occurrence of the P300 in the same timewindow renders the independent assessment of the N400 amplitude difficult. The bulk of the effect of relatedness is typically much later (see Domahs et al, 2007, for an example of a late relatedness effect). As such, the critical measure for sensitivity to the factor of relatedness is likely to be in comparisons of the size (or presence) of this late positivity in each hemisphere.

Our findings will add important new knowledge about how each hemisphere represents mathematical knowledge and uses that information during online processing. If it is the case that both hemispheres are equally sensitive to answer correctness, then that would imply that the poorly performing right hemisphere (in patients with left hemisphere damage) is able to recognize the correct answer at some level, but then might be experiencing a bottleneck in communicating that knowledge externally (in contrast with the common interpretation that the right hemisphere on its own doesn't contain this information). We believe this outcome is possible due to the large body of evidence that both hemispheres are sensitive to expectancy in language – albeit on a different ERP component.

Also of great interest will be the pattern of sensitivity to answer relatedness. A finding that the left hemisphere is most sensitive to this variable would be consistent with the idea that multiplication facts are stored in a left-lateralized verbal memory store. However, given that this sensitivity in central vision has largely been seen on a late positive component rather than on the N400 (c.f., Niedeggen and Rosler, 1999, whose difference wave plots leave the exact componentry ambiguous), the direct linking of semantic memory, the left hemisphere, and relatedness seems unlikely in this paradigm. Instead, a late relatedness effect in the left hemisphere might imply that the left hemisphere fully processes multiple levels of provided answers when assessing them. Interestingly, given that the right hemisphere has been linked with more algorithmic processing of arithmetic problems, it also seems possible that the right hemisphere could show relatedness effects. In combination with a lack of an effect for relatedness in the left hemisphere, such a finding would indicate that the right hemisphere plays a particularly important (and, as yet, underspecified) role in thoroughly analyzing answers.

Our use of ERPs in this study affords us the ability to examine how each hemisphere apprehends and responds to answer types in real time, which will importantly permit us to make novel inferences and construct a broader understanding of hemispheric differences in the domain of arithmetic.

Methods

Participants

Data were analyzed from 30 young adults who were recruited from the University of Illinois community and compensated with either course credit or cash ($10/hour). Participants consented in accordance with the Institutional Review Board of the University of Illinois at Urbana-Champaign. All participants were right-handed native monolingual English speakers (15 female, 15 male) with normal or corrected-to-normal vision, and no early exposure to a second language, history of brain trauma, or current use of psychoactive medications. Their mean age was 20.3 years old (Range: 18-23), and their mean laterality quotient on the Edinburgh handedness inventory (Oldfield, 1971) was 0.85 (Range: 0.45 - 1.00, with 1.0 indicating strongly right-handed and -1.0 strongly left-handed). Nine participants reported having a left-handed biological family member.

In order to reach our target of 30 participants, we collected data from a total of 33 participants because three had to be dropped prior to analysis due to excessive noise either in the global recording or in an individual channel.

Stimuli

Participants saw familiar multiplication problems presented sequentially on the screen (e.g., 3 … x … 4 … = … 12). All 60 critical multiplication problem stems had single digit operands. The only single digit operations that were not included were ones for which creation of all the necessary answer types was not feasible due to how small the operand numbers were: 2 x 3 (and its counterpart 3 x 2), and 2 x 4 (and its counterpart 4 x 2). Square operations (e.g., 6 x 6) were included.

Answer types were either Correct, Incorrect Unrelated, or Incorrect Related. Since Correct answers are naturally distributed to be even numbers about 75% of the time (odd answers are only the result of an odd number times another odd number), both types of Incorrect answers were forced to match this naturally-existing distribution of even numbers. Thus, critical Correct answers were 73% even, Incorrect Unrelated answers were 72% even, and Incorrect Related answers were 75% even. The average answer size participants encountered was about 36 (mean = 36.35), including all Correct/Incorrect answer types.

Incorrect Unrelated answers were generated by taking the Correct answer and either adding and subtracting a number between 1 and 5 to it (with rare cases of 6 when 5 or less generated related answers), thus allowing the creation of Incorrect Unrelated answers that were even in the desired distribution and also that deviated from the Correct answer in an unpredictable manner. Incorrect Unrelated answers were also deliberately selected so as not to accidentally be a correct answer to the sum of the operands (e.g., 5 x 2 = 7 was not permitted as an Incorrect Unrelated answer).

Incorrect Related answers were always a multiple of the first operand, generated by multiplying them with +/- 1 of the second operand. For stems with 9 as the second operand, 8 was used to generate the Incorrect Related answers (since multiples of 10 might be perceived as implausible), and similarly, when 2 was the second operand, Incorrect Related answers were generated by multiplying by 3, not 1.

Half of each type of Incorrect answer were numerically larger than the problem's corresponding Correct answer and the other half were numerically smaller than the Correct answer. On average, Incorrect Related answers were numerically more distant from the Correct answer than were Incorrect Unrelated answers. This is because any time the first operand was above 5, the Incorrect Related answers would necessarily deviate from the Correct answer by a multiple of 6-9, whereas Incorrect Unrelated answers were almost always limited to deviations of 5 from the Correct answer.

There were additionally 30 filler multiplication problems that used 1 as an operand and that used common double-digit operands in very simple problems (e.g., 20 x 3 = 60; 2 x 13 = 26). The answer to these problems was always correct, such that, from the perspective of the participant, the distribution of Correct/Incorrect answers throughout the experiment was 50/50[1]. Due to their relative simplicity, these problems also functioned as intermittent confidence-

[1] The P300 is sensitive to both subjective and global probability, with larger P300s for less probable target items. A 50/50 probability design here therefore reduces the likelihood of observing particularly large target P300s, but even in 50/50 manipulations, targets elicit larger P300s than non-targets (see Johnson, 1985, for factors influencing P300 amplitude).

boosters so that participants would not be discouraged after encountering larger multiples of 7, 8 or 9 (for example).

For the 3 Visual Field presentation conditions (LVF/RH; RVF/LH; Central), there were 3 separate sets of 30 critical items – one set for each of the 3 answer types. The 60 critical problem stems were divided such that there were no exact matches in any group of 30. There could be rare cases of operand order swaps (e.g., a set of 30 could have both 3 x 5 and 5 x 3).

Because the responses were delayed to prevent artifacts during collection of critical data, and because we wanted to prevent early button-pressing, only a subset of trials were selected for participants to respond to. All 90 filler trials were probed, and a subset of the critical trials was probed (33% of the incorrect answers, evenly distributed between unrelated and related conditions across visual fields). From the participants' perspective, just under half of trials were queried (42% total), and 60% were correct.

Six lists were generated so that each combination of answer type (Correct, Incorrect Unrelated, Incorrect Related) and level of VF of presentation (RVF/LH, LVF/RH, Center) was rotated. For example, on List 1, the 30 problem stems in set 3 might have assigned to have Incorrect Unrelated answers in LVF/RH, whereas on List 2, set 3 might have been assigned to have Correct answers in Central vision. Across each set, the distribution of even/odd answers and the distribution of Incorrect answers that were larger/smaller than the Correct answer were examined so as to be matched as closely as possible (although having full rotation of the sets minimized concerns at this level).

Across lists, the order of the 360 problem stems (270 critical: 3 x 3 x 30) was the same.

Procedure

Equations were presented one element at a time on the screen (Operand 1 … x … Operand 2 … = … Answer). All equation contexts were presented centrally, whereas the answers could appear either lateralized or in central vision. Each element was presented for 200ms, then was off for 300ms (SOA of 500ms). After the answers were presented, there was a 1000ms blank period for ERP collection, during which participants were instructed to refrain from blinking or moving their eyes. After the blank, a subset of trials was queried for correctness. Queries occurred after the answer was presented in any of the visual fields in an unpredictable manner. Since queries were postponed, speed was not emphasized to participants.

After the response, if probed, a series of white "~~~~" were presented for 1200ms, during which subjects could blink and move their eyes around. (If there was no probe, this blinking period directly followed the blank period.) A series of red "~~~~" directly followed the blinking period, serving as a warning that the next critical stimulus was about to appear. This warning signal was presented for an unpredictable amount of time ranging from 400ms to 600ms. A small square was continuously presented in the center of the screen just below the stimuli to remind participants where the center was; many participants reported focusing on the dot to keep their eyes steady when the lateralized items appeared. There were periodic self-paced breaks in which yellow "~~~~" were presented after a response – this indicated that the subjects could take a break until they were ready to do more trials. We find that some participants want to use breaks, whereas others find them unnecessary; having them self-paced resolves this problem.

ERP Recording

EEG was recorded using 26 silver/silver-chloride electrodes mounted in an elastic cap. All electrode impedances were kept below 5kΩ, and electrodes were evenly distributed over the scalp (see Figure 2.1 for the arrangement). The data were referenced on-line to the left mastoid and re-referenced offline to an average of the right and left mastoids. A separate frontal electrode acted as ground. The vertical electrooculogram (EOG) signal was monitored with an electrode just below the eye (on the infraorbital ridge) and horizontal EOG was monitored with electrodes placed on the outer canthus of each eye. The EEG signal was sampled at 250 Hz and was subjected to an analog bandpass of 0.02 to 100 Hz during online amplification by Sensorium amplifiers.

Raw waveforms were assessed for inclusion on a trial-by-trial basis with artifact thresholds separately calibrated by visual inspection for each subject. Trials were excluded from averaging if they included blinks, movement artifacts, signal drift, blocking, or a horizontal eye gaze movement. Because participants were only queried following some (but not all) trials, all trials were included in the analysis regardless of subsequent accuracy.[2] For five subjects who had at least 20 blink-contaminated trials, the trials that were excluded from inclusion only due to blink artifacts were corrected and reintroduced to the average (in a procedure described by Dale, 1994). After artifact rejection, critical bins included 26 trials on average, and no individual subject had fewer than 16 critical trials in any bin.

Epochs of EEG data for each trial were taken from 100ms prior to item onset until 920ms after item onset, and the baseline acquired over the 100ms preceding the onset of each trial was subtracted prior to averaging. ERP mean amplitudes were measured after application of a digital bandpass filter of 0.1 to 20 Hz. In all of the results below, interactions and effects of electrode-related factors are only reported for those of theoretical relevance. For all F tests reported with

[2] Trials are not removed on the basis of accuracy in language comprehension / sentence reading tasks either. It is typical in paradigms analyzing expression completions for participants' confidence and interpretation to vary across individual trials.

more than 1 degree of freedom in the numerator, reported p-values are Huynh-Feldt corrected for violations of sphericity.

Arithmetic fluency measure

Following the ERP experiment, participants took a 60-second paper and pencil multiplication test in which they provided answers to as many of the critical single-digit problems in the study as they could. The 60 problems were in the same scrambled order for all participants. In order to minimize test anxiety, this was always described as an "exercise" and never as a test, quiz, or examination.

Results

Behavioral data

On the multiplication fluency test, participants on average produced 38 correct answers for the 60 problems, which is a speed of about 1.5 seconds per written answer (range: 18-60 correct answers). Most participants did not finish in the allotted time, leaving many problems at the end blank (problems were randomly ordered, with easier and harder single-digit problems mixed). The accuracy of reported answers was 98% (most participants, 20/30, did not write any wrong answers). Thus, results of the test found that participants' multiplication fluency was high enough not to be a concern for the validity of interpreting the ERP manipulations, and that, as one would expect, there was some (minor) natural variability in multiplication fact retrieval ability across individuals.

Event-related potentials

General waveform morphology

In Figure 2.2, one can observe the success of the manipulation of visual field. By plotting the lateral occipital channels, effect of visual stimulation in the contralateral visual field is apparent on the N100 component. In the right lateral occipital channel, the visually-evoked N100 is larger for left visual field presentation, reflecting RH stimulation from the contralateral left visual field, and the same is true of N100 in the left lateral occipital channel.

In Figure 2.3, one can observe over posterior channels that a visually evoked P200 component is generated in response to onset of the critical item, with an average latency of 235ms. This effect is almost immediately followed by another positivity, the P300, which is fairly punctate and is over by 400ms post-stimulus. Following the P300, there is a late effect that appears smoother and peaks around 500ms post-stimulus onset.

P300 for Correct Answers

Latency analysis of the P300 effect over all channels found that its peak was 311ms (u = 310.87ms, SE 0.62) after presentation of the critical answer stimulus, with no effect of Visual Field of presentation ($F(2,58) = 0.73$, $p > 0.1$) due to remarkably similar timecourses (RH/LVF peak at 308.56ms, SE = 1.14; LH/RVF peak at 310.15ms, SE = 1.10, Central presentation peak at 313.90ms, SE = 0.96).

In order to characterize its distribution, we ran an omnibus ANOVA on mean values centered around the effect, from 250 to 400ms, with 3 levels of Visual Field (Right VF/Left Hemisphere, Left VF/Right Hemisphere, and Central Presentation), 3 levels of Condition (Correct, Incorrect Related, Incorrect Unrelated), 3 levels of Laterality (right lateral, central midline, left lateral), and 4 levels of Anteriority (from frontal to posterior channels).

This identified a main effect of Condition ($F(2,58) = 9.11$, $p < 0.001$), and interactions between Condition and Laterality ($F(4,116) = 3.38$, $p < 0.05$), Condition and Anteriority ($F(6,174) = 3.90$, $p < 0.05$), and Condition, Anteriority, and Laterality ($F(12,348) = 5.20$, $p < 0.05$). The main effect of Condition was due to larger P300s in all hemispheres to Correct answers (4.49µV, SE = 0.10) relative to Incorrect Unrelated answers (3.77µV, SE = 0.10) and relative to Incorrect Related (3.64µV, SE = 0.10) answers. The interactions between Condition and distributional factors reflected that the P300 effect, while widespread, was largest in centromedial channels, and was quite small in frontal channels (see Figures 2.5 and 2.6 for topographical maps of the distribution). For example, there was no difference between Correct (3.77µV, SE = 0.17) and Incorrect Unrelated, (3.66µV, SE = 0.17) answers in frontal channels, but the mean amplitude of Correct (4.74µV, SE = 0.19) and Incorrect Unrelated (3.75µV, SE = 0.20) differed over middle and posterior channels.

Follow-up comparisons were performed to examine Correctness (Correct compared to Incorrect Unrelated answers) and Relatedness (Incorrect Related compared with Incorrect Unrelated answers) separately, in two ANOVAs including 3 levels of VF, over 21 channels that captured the effect (i.e., five frontal channels were dropped). These tests confirmed that there was a significant effect of Correctness ($F(1,29)= 14.39$, $p < 0.001$) and no effect of Relatedness ($F(1,29) = 0.04$, $p > 0.5$), nor of Visual Field in either test ($F(2,58)$s < 0.32, $p > 0.5$).

Overall effect of Late Positivity

In order to characterize the distribution of the late positive effect, we began by running an omnibus ANOVA with 3 levels of Visual Field (Right VF, Left VF, and Central Presentation), 3 levels of Condition (Correct, Incorrect Related, Incorrect Unrelated), 3 levels of Laterality (right lateral, central midline, left lateral), and 4 levels of Anteriority (from frontal to posterior channels) over the time period 400-600ms following presentation of the critical answer stimulus. For all F tests reported with more than 1 degree of freedom in the numerator, reported p-values are Huynh-Feldt corrected for potential violations of sphericity.

This revealed main effects of Visual Field ($F(2,58) = 4.44$, $p < 0.05$), Condition ($F(2,58) = 9.32$, $p < 0.001$), Laterality ($F(2,58) = 49.89$, $p < 0.001$), and Anteriority ($F(3,87) = 34.75$, $p < 0.001$).

Additionally, there was a significant four-way interaction between all factors (Visual Field, Condition, Anteriority, and Laterality), ($F(24,696) = 1.8$, $p < 0.05$), reflecting that effects of Condition between visual fields were distributed unevenly across the scalp. Specifically, when Condition-related effects were present in a VF, they were largest over central and posterior sites, and reduced or absent in other locations (e.g., in LVF/RH presentation, there was a 0.79μV difference between Correct (0.37μV, SE = 0.50) and Incorrect Unrelated (1.15μV, SE = 0.42) in a representative frontal channel, whereas in a representative posterior channel there was a 1.93μV difference between Correct (3.03μV, SE = 0.43) and Incorrect Unrelated (4.96μV, SE = 0.47)).

In order to test prior planned comparisons for the effect of Correctness (Correct versus Incorrect Unrelated) and Relatedness (Incorrect Related versus Incorrect Unrelated) in each hemisphere/VF on the LPC, a cluster of 13 electrodes from the peak of the effect (as indicated by the distributional analysis, see Figure 2.6) is therefore used in the following analyses.

Central Presentation

There was a significant effect of Correctness: (($F(1,29) = 4.67$, $p < 0.05$) due to more positivity in response to Incorrect Unrelated answers (4.25μV, SE = 0.16) than to Correct answers (3.21μV, SE = 0.12). In testing the effect of Relatedness, although numerically Incorrect Unrelated answers elicited a more positive response (4.25μV, SE = 0.16) than did Incorrect Related answers (3.94μV, SE = 0.15), this was not significant ($F(1,29) = 0.57$, $p > 0.1$).

Right Hemisphere

There was both an effect of Correctness ($F(1,29) = 9.68$, $p < 0.01$) and Relatedness ($F(1,29) = 7.94$, $p < 0.01$) . This was driven by significantly more positive responses to Incorrect Unrelated answers (4.71μV, SE = 0.14) compared to both Correct answers (3.24μV, SE = 0.14) and Incorrect Related answers (3.65μV, SE = 0.13). Numerically and visually (see Figure 2.4), the effect of Correctness was larger than Relatedness, although a post-hoc follow-up comparison of Incorrect Related and Correct answers found no difference between these conditions in this timewindow ($F(1,29) = 1.64$, $p > 0.21$).

Left Hemisphere

There were no effects of Correctness ($F(1,29) = 2.46$, $p > 0.1$) or Relatedness ($F(1,29) = 0.54$, $p > 0.1$) in this timewindow following RVF/LH presentation. The absence of effects as compared to LVF/RH can be seen in Figures 2.4 and 2.6.

Discussion

Our data from central presentation generally replicates what has been previously shown in prior papers using these manipulations with arithmetic expressions, despite moderate changes to the precise stimulus design (e.g., our incorrect answers of both types were generated by different means than has been used before). Specifically, we replicate the presence of an early distinction between correct and incorrect answers in the form of a larger P300 to correct answers (although this same effect has a history of being labeled an N400 by others). Importantly, the actual waveform morphology of a positivity to correct answers is the same across these studies (see Jasinski & Coch, 2012). In the broader literature, the P300 is enhanced to target items, even when probability is held constant, and its latency is generally taken as a measure of stimulus evaluation time. The larger P300 to correct items, peaking around 300ms in this experiment, implies that these items were identified and categorized as the target item quite quickly, likely because of there was a full second between onset of the second operand and onset of the final answer. In cases with reduced time to predict and expect the final target answer, the P300 has been delayed in time – hence occurring more in the traditional N400 timewindow -- and/or reduced in amplitude (indeed, if at all present, e.g., Domahs et al, 2007, SOA 150ms; Niedeggen, Rösler, & Jost, 1999, SOA 200ms). Our result in central presentation largely implies that the system had sufficient time to process the problem prior to presentation of the answer.

That we did not find support for a relatedness effect in central vision suggests that the generation of this effect is relatively more sensitive to nuances of stimulus design. The timing of presentation is one potential reason for a lack of a significant late positivity – effects of relatedness might be smaller when the system has more time to fully process the problem and retrieve the correct answer. Alternatively, in preparation of our incorrect unrelated answers, we carefully matched the distribution of even and odd numbers in our related and unrelated incorrect answers to the distribution of correct answers (because they are more likely to be even). This is not a process that we have seen others regularly do, leaving open the possibility that sensitivity to relatedness might be partially a sensitivity to those answers being more like correct answers on the dimension of evenness, compared to incorrect unrelated answers (which are more likely to be odd unless intervention occurs). Given that the P300 is sensitive to subjective probability, it is possible that the relative evenness and oddness of answers across conditions could be influencing its amplitude in these studies, although the result of direct manipulation of these variables has not yet been examined.

Our primary aim was to characterize how the effects of correctness and relatedness might differ as a function of hemisphere, given the prevalence of differences across patient groups with unilateral lesions, (Rosselli & Ardila, 1989; Cohen et al, 1999; for review, see Cappelletti & Cipolotti, 2010) and given the effects of lateralized presentation in patients who have undergone commissurotomies (Gazzaniga & Smylie, 1984; Colvin, Funnell, & Gazzaniga, 2005; Funnell et al, 2007). Through examining these effects, we are able to address two fundamental questions: 1) How well is each hemisphere able to recognize and accurately categorize correct answers from incorrect answers? and 2) How sensitive is each hemisphere to the putative table-based organization of knowledge about mathematical facts, as primed here by different problem contexts? Interpretation and discussion of our results follows for each issue in turn, with consideration given to the corresponding findings in sentence processing.

We found that both hemispheres were able to recognize and categorize the correct answer as correct (indicated by a larger P300 to this condition than to incorrect answers), and were able to do so at the same speed (indicated by no latency difference in the P300 across visual field presentations). In order to reconcile this with prior reports that the right hemisphere is less critical for production of answers than is the left (i.e., patients with left hemisphere lesions are, consistently across studies, more likely to have difficulties with arithmetic), it is important to remember that our recordings do not reflect the hemisphere's abilities to output their knowledge in the form of a response (verbal or otherwise) – instead, our ERPs reflect each hemisphere's appreciation for the answer types in the absence of any form of overt response (which was only requested after a substantial delay). It is well documented that the P300 specifically indexes stimulus evaluation time, and not response selection and planning/execution, which are otherwise difficulty to disentangle in reaction time measures (Kutas, McCarthy & Donchin, 1977). Our findings are consistent with the interpretation that the hemispheres are able to recognize correct from incorrect answers at the same timescale, but subsequently undergo different processes when selecting and executing outward expressions of this internal categorization process.

This pattern aligns with typical results in studies of language comprehension, in which participants passively read sentences and that end in expected or unexpected ways. Both hemispheres have been shown to able to distinguish "correct" final words from "incorrect" final words (and consistent / inconsistent line drawing images as well), albeit on a different ERP component (the N400, Coulson, Federmeier, Van Petten, & Kutas, 2005; Federmeier, Mai, & Kutas, 2005; Wlotko & Federmeier, 2007). The similarity in response patterns across hemispheres in those studies has been taken as evidence that both hemispheres are able to make sense of the provided sentential context and integrate new words into the message. Our findings here are similarly consistent with the idea that both hemispheres are able to take the context provided by mathematical expressions (multiplication problems) and recognize whether provided answers are good fits or poor fits, although ultimately they might not be equally able to overtly communicate that knowledge.

That this main effect of congruency occurs on different components across language and arithmetic studies is intriguing. It is perhaps understandable why in sentence processing studies there are typically no reported P300 effects. The goal in these experiments tends to be to "read for understanding", not to read and make an explicit judgment specifically about the correctness of the critical (typically sentence-final) word. When people do make explicit, categorical judgments about words in a sentence (or at least about the sentence as a whole) then the critical words that typically contain information relevant for making the categorical judgment often elicit positivities as well (see, e.g., Geyer, Holcomb, Kuperberg, & Perlmutter, 2006, and Kim & Osterhout, 2005, for positivities elicited in the context of plausibility judgments).

What is harder to reconcile is why there is not also an N400 effect of congruency in the context of math equation processing. Symbolic digits presumably convey meaning (in the form of numerosity and other long term memory representations), but perhaps this information is not fully appreciated by the N400 timewindow. We have previously observed that some features of wordforms that have been thought to be part of the meaning of the word only are appreciated later in the form of a late positive complex (LPC), and the same might be true of the nuanced meaning of symbolic numbers (Fischer-Baum, Dickson, & Federmeier, 2014).

On the other hand, it is possible that the P300 effect is overlapping with a co-occurring (smaller) N400 effect, and that the effect of semantic facilitation is present, but hard to detect. Given that all potentially meaningful stimuli elicit N400 components, and that digits fall within that category, we believe this to be possible (Federmeier & Kutas, 2011). However, the effects in the present study are clearly unlike typical N400 effects of semantic facilitation in both morphology (with a distinct positivity for the "facilitated" correct condition) and in timing (occurring almost 100ms prior to the traditional N400 timewindow.) To more clearly examine semantic processing of digits, the experimental design would need to contain tasks that do not also elicit substantial P300 effects.

Regarding sensitivity to subtypes of incorrect answers, we only find evidence for an effect of relatedness in the LVF/RH (in the form of a larger late positivity for incorrect unrelated answers compared to incorrect related answers), and no significant difference between these answer types in either the left hemisphere or in central presentation (although there was a numerical trend for the effect during central viewing). This result provides the first evidence, to our knowledge, that, whereas the right hemisphere appreciates a given incorrect answer's similarity to the correct answer, the left hemisphere does not.

This pattern is a remarkable reversal of what was previously reported for a similar manipulation in sentence processing, in which unexpected sentence-final words that were semantically related to the expected completion were only distinguished from unexpected unrelated words following presentation to the RVF/LH (i.e., the left hemisphere had a relatedness effect) (Federmeier & Kutas, 1999b). In that work, it was instead the right hemisphere that had a more categorical approach in comprehension of completions. However, at a mechanistic level, direct comparisons of these results to those reported here is difficult because, whereas the effect of relatedness in sentence processing was observed on the N400, the effect of relatedness we report is on a late positivity.

Effects in language processing on late positivities are common across studies, although the particular eliciting conditions and nomenclature for the subsequently reported effects varies (see Van Petten & Luka, 2012, for a review of the ongoing differences of opinion and complicated discussions unfolding in this domain). A few studies have looked at late positivities in each of the hemispheres. In one, there was greater late positive complex (LPC) priming for backward ordered word pairs in the left hemisphere (e.g., "meat" followed by "butcher", where "meat" does not strongly prime "butcher", but the words are clearly related) than in the right, which suggests that there is a left hemisphere advantage for strategically reshaping meaning activation for weakly related and/or non-canonically ordered word pairs (Kandhadai & Federmeier, 2010a). In a subsequent publication, it was observed that the right hemisphere could be forced to appreciate the relatedness in these types of word pairs, but, on the whole, there is largely evidence of hemispheric biases in language contexts favoring the *left* hemisphere for sensitivity to relatedness on late positive effects (Kandhadai & Federmeier, 2010b). Our result of a right hemisphere relatedness effect thus serves as strong evidence that the memory system engaged by multiplication arithmetic problems and the memory system engaged in sentence and word reading can be dissociable, and their interrelationship is more complicated than it might appear on the surface.

Because there are currently no published reports of late positivities *selective* to the right hemisphere, and because there are also no studies looking specifically at behavioral responses to these particular conditions with visual half-field presentation, there are many alternative interpretations for why, in the present work, there is a late positivity selective to the right hemisphere that is sensitive to the relatedness of answer types. That is, at a mechanistic level, the source of the right hemisphere appreciation of related answers could be the result of a number of differences in how learned information is structured and/or in how incoming information is processed by each hemisphere. We will therefore review three possible sources of the effect below.

Under a first interpretation, this right hemisphere relatedness sensitivity finding could be because arithmetic fact knowledge is genuinely organized in a table-like fashion in the right hemisphere and not organized in this way in the left hemisphere. Then, when related answers are provided, only the right hemisphere is prepared to recognize this property. Although this is possible, we don't have any evidence to justify why this would be the case. In fact, our finding for similar effects of correctness across the hemispheres is also possibly inconsistent with this view because it might be expected that there should be divergences in responsivity on the P300 during this timewindow if its memory were indeed structured so differently in the hemispheres.

A second possibility is that the left hemisphere also contains the table-relatedness information but does not engage in accessing it. In fact, this might be advantageous: it could be harmful for the successful execution of the task at hand to be entertaining multiple answers after seeing the full arithmetic expression. Under this view, the left hemisphere could be strategically focusing on the search for only the correct answer. Interestingly, this idea is consistent with the proposal that the right hemisphere generates a broader range of concepts than does the left (Jung-Beeman, 2005), which in this case would imply that the right hemisphere activated a broader range of potential answers. If the right hemisphere does this, it might explain some of its apparent difficulty in response selection. This interpretation would effectively render the right hemisphere relatedness effect as reflecting a kind of "confusion", which could serve as a potential explanation for why the isolated right hemispheres in commissurotomized patients performs poorly in verification tasks (e.g., Funnel et al, 2007). Although this interpretation is certainly possible, it is difficult to heavily favor it because multiple direct tests of the coarse coding / spreading activation hypothesis has not found support for it (Kandhadai & Federmeier, 2007, 2008). Also problematic for this view is that one might have expected to see earlier effects of relatedness on the P300 in the right hemisphere, reflecting a lack of focus on the correct answer as a target item, which is not what was observed.

A final mechanistic explanation places the locus of this effect more in the domain of explicit evaluation of the provided answer. This is in contrast to the prior two explanations, which place the source of the difference between the hemispheres response to related answers either at the level of differences in long-term memory structures or in differences in spreading activation following presentation of the problem itself. Under this third view, once the left hemisphere sees an answer, it categorizes it as correct or not (i.e., generating a P300), and then does no more follow-up evaluation the answers. In the right hemisphere, once an answer has been presented for consideration, it could be engaging in a deliberative appreciation for subtleties in properties of provided answers. Although this might be functionally important for fully analyzing arithmetic problems beyond an instinctual recall of correct answers, and having more information about the qualities of the provided answers could in theory be helpful for deliberative mathematical reasoning tasks, it is nonetheless the case that the left hemisphere's theoretically more categorical approach would be less likely to be vulnerable to error production.

In general, this potential interpretation of the relatedness effect in the right hemisphere is also consistent with the idea that the calculating right hemisphere is involved in more nuanced evaluations of approximate values (perhaps drawing on enhanced spatial abilities), whereas the left hemisphere's approach to answer evaluation has been attributed to relatively more simple memorized retrieval of mathematical facts (i.e., recognizing the correct answer as correct) (Dehaene & Cohen, 1995; Rosselli & Ardila, 1989). However, it is important to note that we have evidence that the right hemisphere initially does engage in the categorization process that characterizes the left hemisphere; it is only later in time that the right and left hemispheres' processing strategies diverge and the right could begin this process of appreciating more detailed properties of the proposed answers.

On the whole, we therefore find this final interpretation to be most consistent with the existing literature on both arithmetic and language, and it is the only interpretation of the three we introduced that does not also make a prediction that the P300 in the right hemisphere should have a relatedness effect.

One important factor that has not been discussed thus far is that of the amount of time our participants had to process the full multiplication expression prior to the answer appearing on the screen. This one-second interval between seeing the second operand (i.e., the complete problem) and seeing the answer is not very pressured, especially compared with some paradigms wherein this interval was reduced to 150ms (in particular, in Domahs et al, 2007, this presentation latency visually appears to have substantially delayed and reduced the amplitude of the P300 compared to the when the P300 was elicited with a 550ms stimulus onset asynchrony). Moving forward, then, it could be of interest to see how the hemispheres respond when they are under more pressure (i.e., in speeded response conditions, and/or with less time between viewing the operands and viewing the answers). It is possible that the left hemisphere, which has been associated with rapid top-down processing of input, might be differently affected by such a manipulation than would the right hemisphere, which has been associated with slower, more bottom-up processing (Federmeier, 2007).

This work is the first to characterize the temporal dynamics of answer evaluation processing in the two hemispheres, and, as such, our findings make a novel and important contribution to the growing field of interest in the neurobiology of arithmetic processing. We found that, similar to language, both hemispheres are sensitive to whether or not an expression has been completed in an expected / correct or unexpected / incorrect manner, although the mechanisms supporting this sensitivity seem to be functionally different. Most intriguingly, we uncovered a unique sensitivity to properties answer, beyond their status as correct or incorrect, that occurs later in time only in the right hemisphere, which appears to reflect its more deliberative processing strategy. Given that a similar effect has not yet been observed in language processing, these results are consistent with a view that there are important differences in how we read and assess sentences in language and expressions in arithmetic problems.

Figures

Figure 2.1

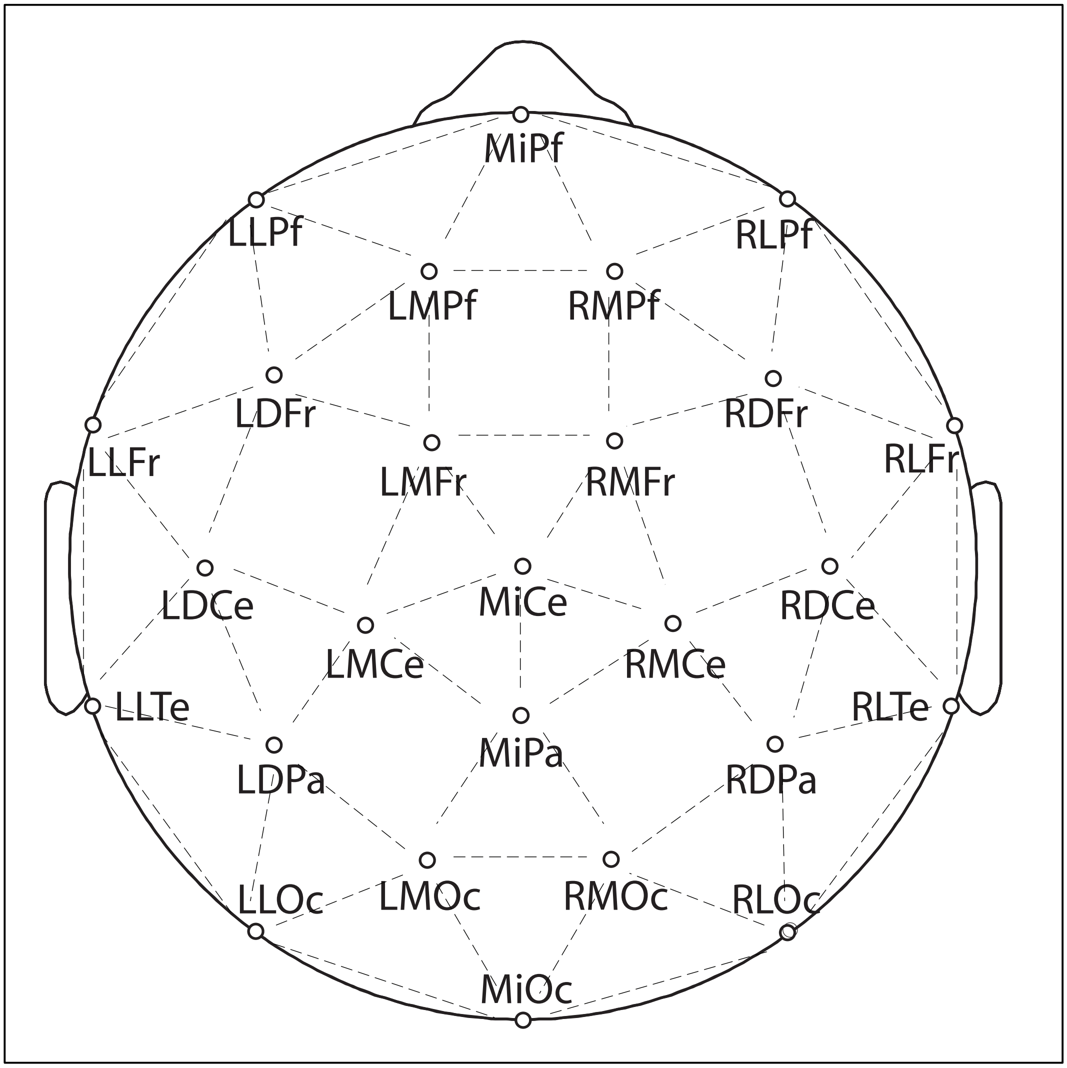

Figure 2.1 A graphical representation of the configuration of the 26 channels used in critical EEG recordings.

Figure 2.2

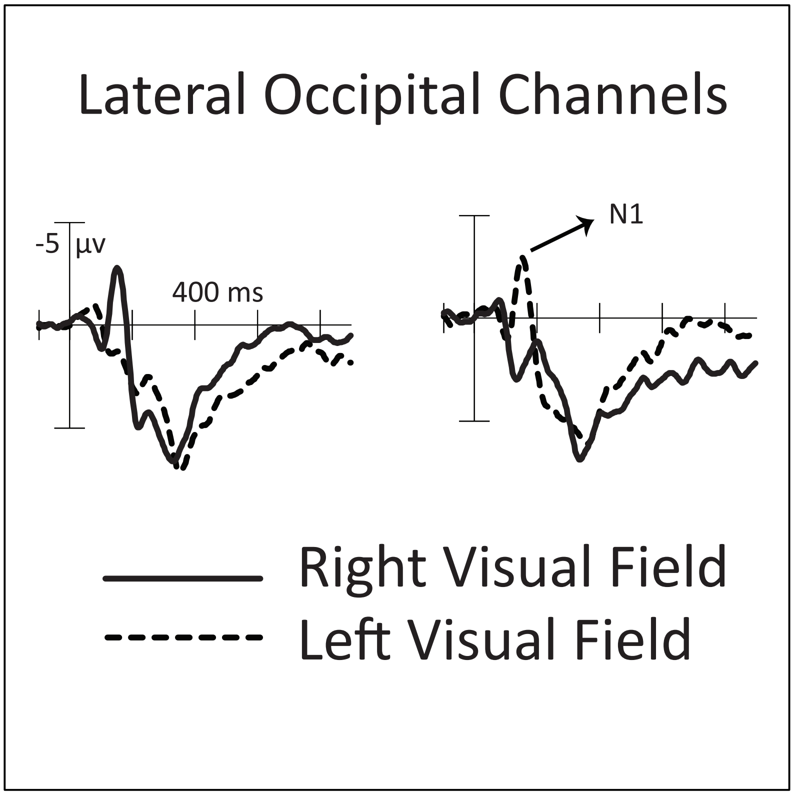

Figure 2.2 This displays the recordings from left and right lateral occipital channels (LLOC at left and RLOC at right) following presentation to the right visual field (biased to the left hemisphere, plotted with solid black), and presentation to the left visual field (biased to the right hemisphere, plotted with dashed black). The N1 visually evoked potential is larger for the scalp channels corresponding to the stimulated hemisphere. There is also a later, slow-wave effect (starting here around 400 ms post-stimulus) that is characteristic of visual field manipulations.

Figure 2.3

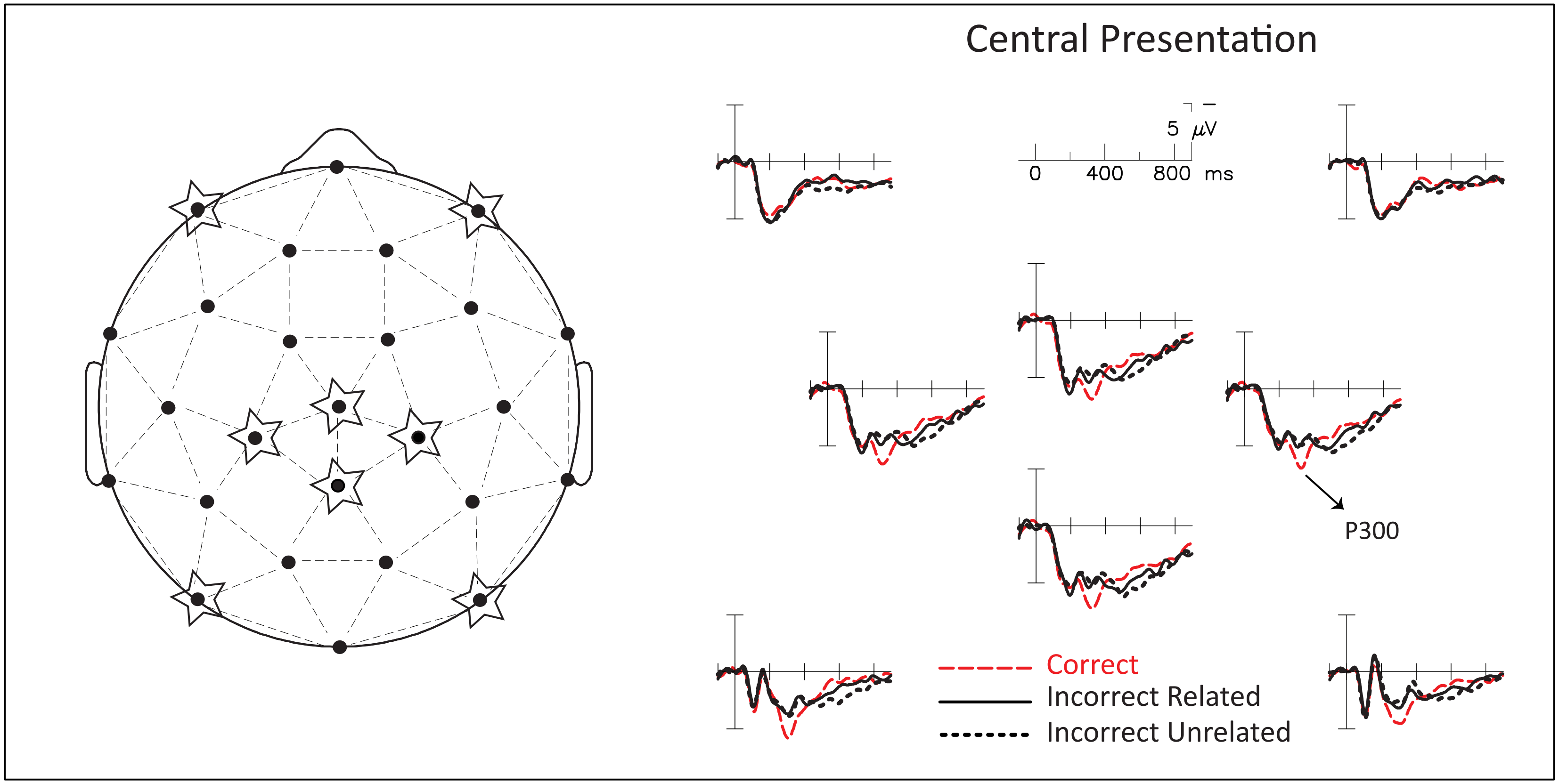

Figure 2.3 Responses to centrally presented critical answer types from selected channels over the head (starred on the channel configuration plot at left) are plotted. The P300 elicited by correct (red) answers, as compared to incorrect related (solid black) and incorrect unrelated (dashed black), can be seen in the centromedial and posterior channels (and not the frontal channels at top).

Figure 2.4

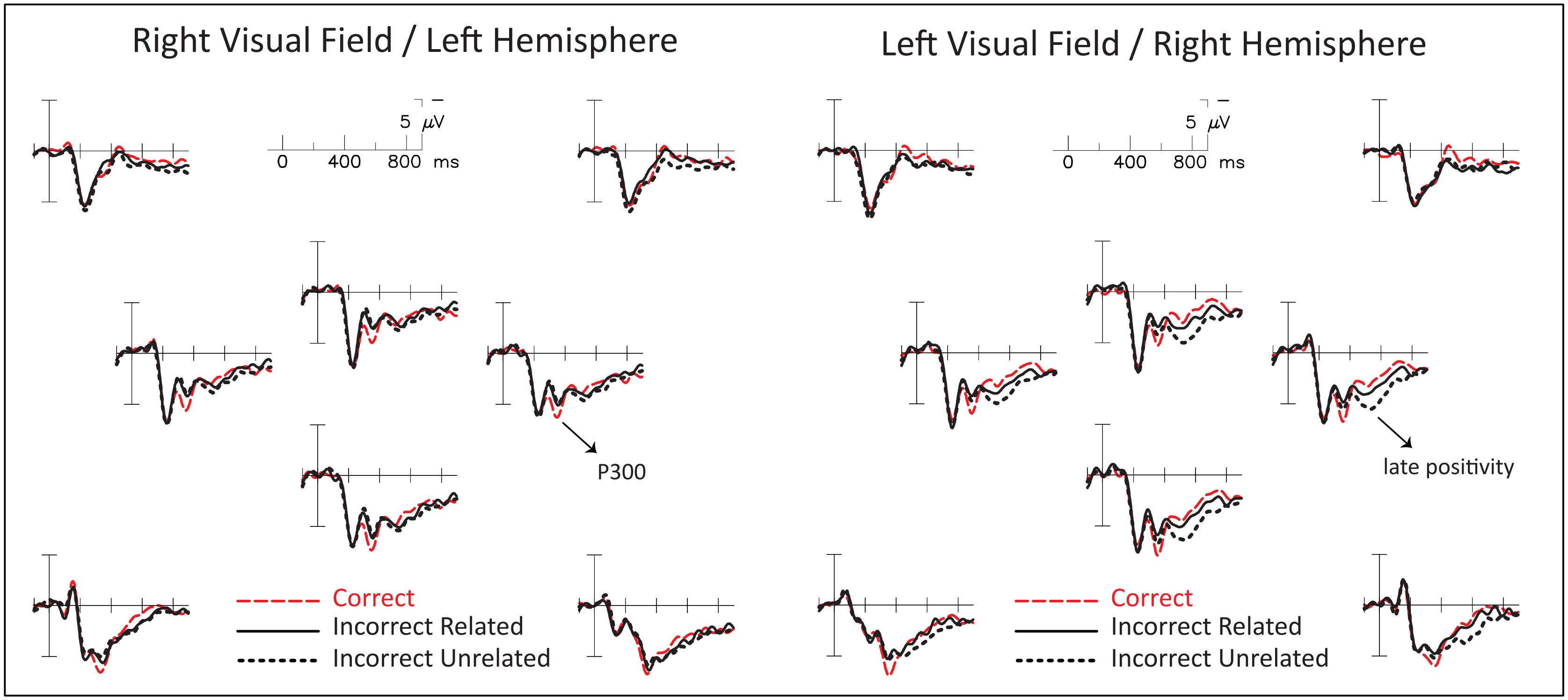

Figure 2.4 Responses to critical answer types after presentation to RVF/LH (left plots) or LVF/RH (right plots) from selected channels over the head (same as in Fig. 1.3) are plotted. The P300 elicited by correct (red) answers, as compared to incorrect related (solid black) and incorrect unrelated (dashed black) can be seen in the centromedial and posterior channels in both LVF/RH and RVF/LH, and the later effect of relatedness can only be seen in the LVF/RH plots (difference between two black lines in central/right channels).

Figure 2.5

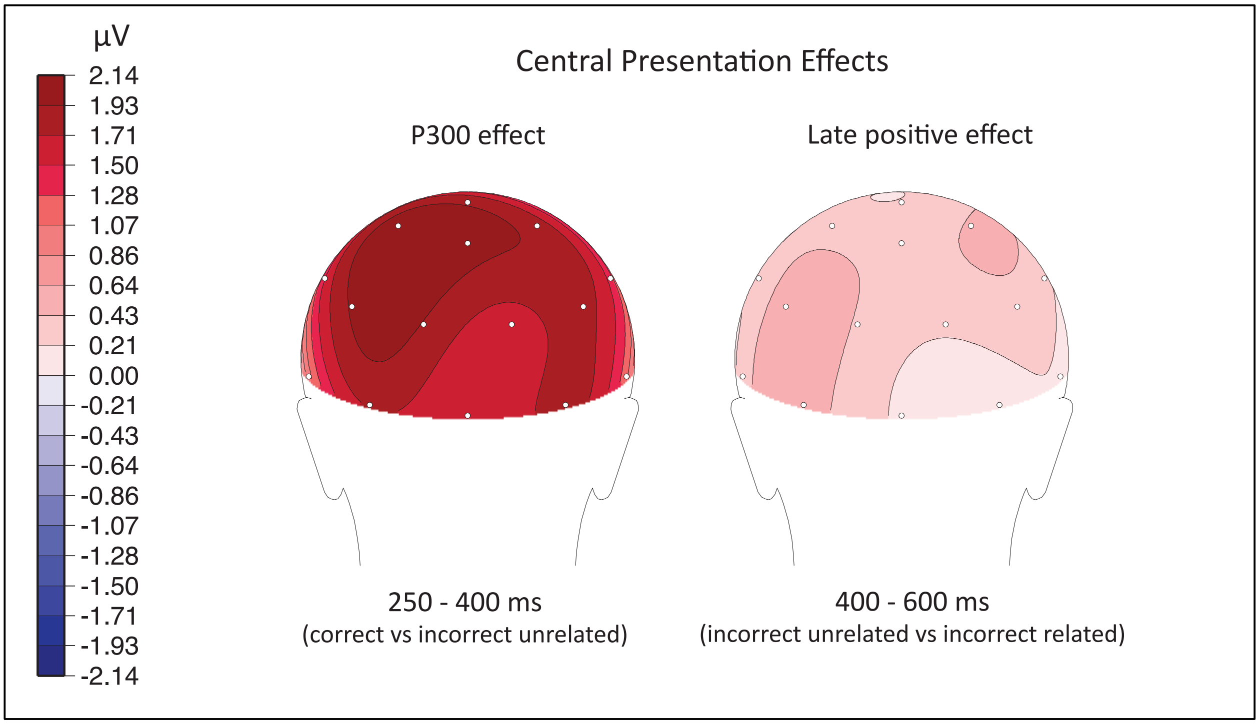

Figure 2.5 Topographical maps of the distribution of the effect of correctness (left) and relatedness (right), as measured by the subtraction of the response to incorrect unrelated answers from correct answers during the P300 timewindow (measured from 250 – 400 ms post stimulus onset), and subtraction of the response to incorrect related answers from incorrect unrelated answers during the later timewindow (measured 400 – 600 ms post stimulus onset), as viewed over the back of the head. Darker areas correspond to more substantial effects. It can be seen that there is an effect of correctness in the P300 timewindow and no effect of relatedness in the later timewindow.

Figure 2.6

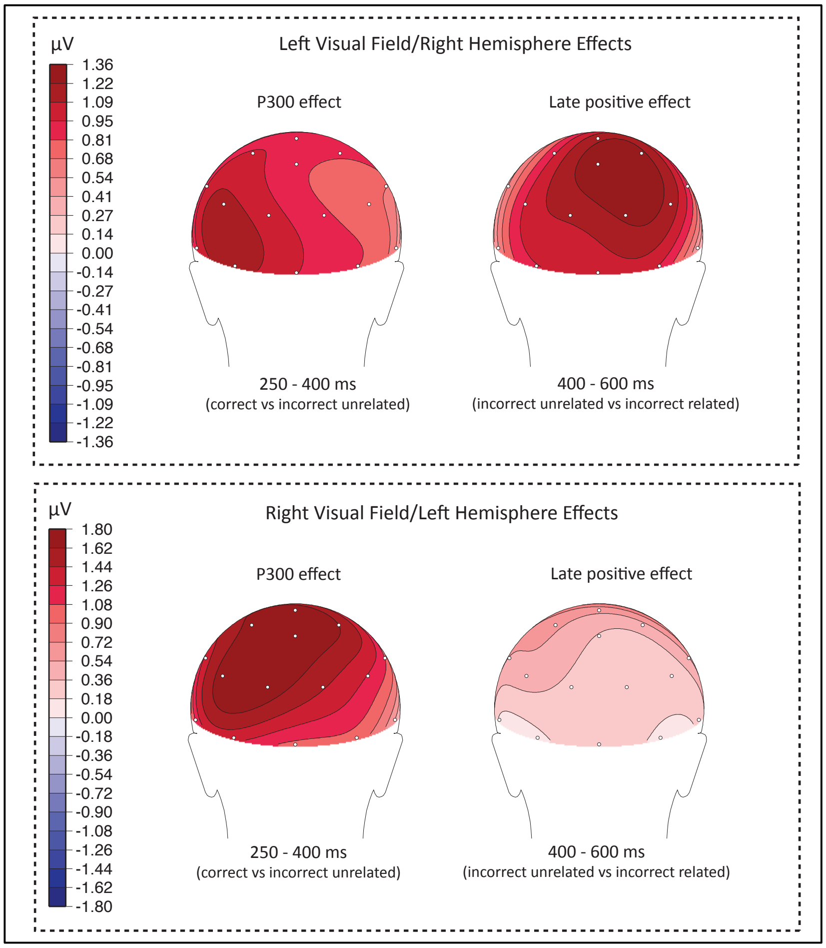

Figure 2.6 Topographical maps of the distribution of the effect of correctness (plots on left) and relatedness (plots on right), as measured by the subtraction of the response to incorrect unrelated answers from correct answers during the P300 timewindow (measured from 250 – 400 ms post stimulus onset), and the response to incorrect related answers from incorrect unrelated answers during the later timewindow (measured 400 – 600 ms post stimulus onset), viewed over the back of the head. Darker areas correspond to more substantial effects. It can be seen that there is an effect of correctness in the P300 timewindow for both RVF/LH (bottom left plot) and LVF/RH (top left plot), but an effect of relatedness only for responses following LVF/RH presentation (top right plot).

CHAPTER 3

Experiment 2: Behavioral effects of answer evaluation across the hemispheres

Introduction

The results from Experiment 1 suggest that there are both similarities and differences in how the hemispheres process multiplication problems. Effect patterns on the P300 revealed similarities in how each hemisphere evaluates solutions to such problems: both hemispheres were sensitive to the overall correctness of answers and made this distinction similarly quickly (as indicated by P300 latency). However, the hemispheres' differential sensitivity to incorrect answer types on a later positivity suggests important differences in how the RH and LH represent numerical relationships and/or process math equations. Thus, ERPs revealed substantial mechanistic processing differences between categorizing an answer as correct versus incorrect, and performing more deliberative and/or prolonged analyses of provided answers.

Little is known about how these differences manifest behaviorally – in fact, there is only one behavioral experiment looking at arithmetic processing across the hemispheres in participants without brain damage (Castro, Sumich, Premkumar, & Jones, 2014). That study, which grouped all arithmetic operation types together, reported a disadvantage for accuracy in responding to mathematical expressions with LVF/RH presentation (compared to both RVF/LH and bilateral presentation), but only in the context of a complicated interaction with answer type and notational consistency. Thus, no study has specifically examined response time and accuracy differences across the hemispheres in the context of multiplication problems. However, the ERP results from Experiment 1 permit some predictions about likely outcomes. The present study is therefore designed to look at the behavioral consequences of the processing differences uncovered in the prior experiment, with the aim of better understanding the nature and impact of hemispheric differences in numerical processing.

In the language domain, it is typical for behavioral indices to favor RVF/LH presentation (faster RTs, higher accuracies) in most basic comprehension tasks, an outcome that has been attributed to the LH's superior word recognition abilities (Jordan, Patching, & Thomas, 2003; Jordan, Patching, & Milner, 2000; see Hellige, 1993 for an overview). However, the P300 effect pattern obtained in Experiment 1 suggests that similar LH advantages may not obtain in math, and, instead, that there might not be any substantial differences across the hemispheres in their response times, at least to correct answers. This prediction can be formed because P300 latency (which was the same across the hemispheres) is highly correlated with reaction time, as reviewed below.

Prior work examining the relationship between P300 latency and RT has indicated that there are strong correlations between the two under many circumstances (e.g., Ritter, Simson, & Vaughan, 1972). Stronger alignment between P300 latencies and RTs are obtained when timing differences arise at the level of stimulus evaluation, and when people are not required to emphasize speed in their responding (which is associated with responding in the context of incomplete stimulus evaluation; e.g., Kutas, McCarthy & Donchin, 1977). Dissociations between these measures have been found to occur when a delay in processing arises at later stages, such as response selection (Ragot & Renault, 1981). For example, whereas RTs are sensitive to manipulations of response compatibility (e.g., needing to respond with a left hand button press to the word "RIGHT"; McCarthy & Donchin, 1981), P300 latencies are not, making them a more selective measure of timing shifts related to stimulus evaluation. In sum, given the findings in Experiment 1, and the relationship between RTs and P300 latency, VF-based differences in RT would suggest hemispheric differences that arise after the stimulus evaluation processes that were reflected in the P300.

Interestingly, the (initial) P300 in response to incorrect answers occurred at the same time as the target P300 for correct answers (albeit being smaller in amplitude), suggesting that there also might not be an RT delay for responding to incorrect answers. However, the behavioral arithmetic literature strongly suggests that we should observe slower RTs to incorrect compared to correct answers (e.g., Lemaire, Abdi, & Fayol, 1996). It has been shown in the broader response time literature that, although verification is typically faster than falsification, speed of yes/no matching judgments can be changed as a function of how cautious participants are told to be (i.e., if told to be more cautious about responding “yes”, participants can even become slower to verify than to respond “no”), and that, more generally, differences in criterion-setting can explain RT delays for “no” responses, without needing to invoke additional processes to explain these delays (Ratcliff & Hacker, 1981; Ratcliff & Rouder, 1998). Given that our instructions are the same as is generally used in the arithmetic verification domain (and thus that we are not introducing new or different response biases), we expect to obtain an RT delay for incorrect trials despite the similarity in the P300 latencies for these conditions.

In addition to delays for incorrect trials as a whole, incorrect answers that are related to the correct answers are associated with even slower RTs. Through dual-task manipulations and studies of special populations, it has been revealed that there are likely additional cognitive processes (e.g., inhibition and working memory resources) that are specifically engaged when rejecting these provided solutions (e.g., Barrouillet, Fayol, & Lathulière, 1997; Rammelaere, Stuyven, & Vandierendonck, 1999; Lemaire, Adbi, & Fayol, 1996). In the ERP data as well, differences between unrelated and related incorrect answers emerged in the post-P300 time window, in the form of differential amplitude of a component in the Late Positive Complex (LPC) family. Therefore, although we, again, did not obtain latency differences on the initial P300 response to these items compared to other answer types, we expected to replicate the finding that RTs to these items are slowed, possibly because of differential processing that takes place after initial stimulus evaluation.

In general, then, there are two main effects that are expected to lead to RT delays – one being answer correctness (with slower RTs for incorrect answers), and the second being incorrect answer relatedness (with the slowest RTs for incorrect related answers); the second of these in particular may be associated with unique processing demands (as evidenced by both our prior ERP results and prior patterns in the behavioral literature). Our goal, then, is to further examine these anticipated RT patterns in the context of visual field manipulations that offer the opportunity to link hemispheric processing differences to behavioral outcomes.

One potential approach to examining the impact of these processing differences at a behavioral level would be to measure mean reaction times across our visual field and answer type conditions. Indeed, measurements of the mean are the traditional source of reported RT effects in this domain (e.g., Ashcraft, 1992; Lemaire & Reder, 1999; Campbell & Tarling, 1996; Zbradoff & Logan, 1990). However, there is another method of examining RTs that seems more readily capable of capturing the potential distinction between the processes associated with effects of correctness and effects of relatedness (which were associated with different brain responses). In literatures outside the arithmetic domain (e.g., language, aging), there have been many reports that experimental manipulations not only affect mean RTs, but also alter the underlying shape of the distribution at a trial-by-trial level (see Balota, et al., 2008, for review).

One of the key features of the relatedness effect in particular is its characteristic association with very slow RTs, resulting from the additional processing mechanisms required to reject these easily confusable answers. This is likely to manifest in the form of relatively more RTs in the tail of the distribution as compared to just a uniform shift in the peak of the mean to the right (e.g., Staub & Benatar, 2014, for an example of this). In order to specifically address this possibility, we opted to fit our data to the ex-Gaussian distribution, which has parameters that can separately capture overall shifts of the distribution as a whole from alterations to the shape of the distribution in the form of rightward skewing (i.e., disproportionately very slow RTs).

This distribution is a convolution of the Gaussian and exponential probability distributions (yielding its name), and its shape is defined by three parameters: μ, σ, and τ (reviewed in Balota & Yap, 2011). Two parameters (μ and σ) are derived from the Gaussian component of the distribution: the modal central tendency is reflected in μ, and variability in this modal portion of the distribution is captured by σ, the standard deviation. Critically, these parameters are not as influenced by rightward skews as measurements of the mean are. Instead, the rate parameter, τ, from the exponential component of the distribution, captures the amount of rightward pull in the tail of the distribution due to slower RTs. An additional feature of using this distribution is that its parameters are readily linked back to tradition RT measures: the sum of μ and τ correspond to an approximation of the mean, such that, if there *is* no disproportionate rightward distribution of RTs, then τ would be 0, and μ would simply correspond to the “normal” mean (see Fig. 1 of Balota & Yap).

With respect to the current study, we anticipate that effects of correctness and effects of relatedness might manifest differently on μ and τ – and, specifically, that there might be a difference on τ that is unique to rejecting incorrect but related answers. This would be particularly interesting given that increases in τ manifest on more challenging decision-making processes and have been linked to increased response competition (Balota & Spieler, 1999; McVay & Kane, 2012; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). There is also a growing field measuring individual differences in properties of RT distributions, which have correlated higher τ with lower working memory span. This effect has been interpreted under the view that people with lower working memory abilities also have a disproportionate number of trials in the slow tail of their RT distribution as compared with higher-span individuals whose responses cluster more closely around the mode (e.g., Tse, et al., 2010; Payne & Stine-Morrow, 2014). Alternatively, there is an account of τ that associates its increases with a disruption in normal, ongoing processing (Staub & Benatar, 2014), which might be expected to occur when participants encounter either type of incorrect answers relative to correct. In sum, we may be able use this property of the RT distribution (τ) to make inferential links between arithmetic processing and the broader decision-making field.

Of particular interest, of course, will be whether any of these effects manifest differently across the VFs. A straightforward mapping of effect sensitivity in each hemisphere would predict that, since only the RH showed an effect of relatedness at the level of brain activity, the expected RT delay should emerge for LVF/RH but not RVF/LH processing. However, there are theories of interhemispheric communication that suggest that the RH is particularly fast at sharing its knowledge with the LH, which would instead predict that relatedness effects from the RH might also be able to manifest behaviorally under RVF/LH processing conditions (e.g., Marzi, Bisiacchi, & Nicoletti, 1991; Barnett & Corballis, 2005). Moreover, the particular demands of making manual responses with both hands may entail a level of cooperation that would mean that the effect of relatedness emerges in the RVF/LH case as well. That is, because motor planning in both hemispheres is required to execute the necessary manual responses (and yes/no response mapping is counterbalanced with response hand across participants), decision-making information that originates from the LVF/RH might necessarily spread to the LH and manifest behaviorally. It might not be surprising, then, if the relatedness effects that emerged only under LVF/RH presentation conditions in brain activity would be more widespread in behavioral testing.

In sum, in Experiment 2 we compared how effects of correctness and relatedness manifest across parameters that emphasize different features of the underlying RT distribution (i.e., the location of the modal response versus the amount of their rightward skew), which may provide further evidence that the different processes that were revealed in ERPs also are associated with mechanistically different behavioral responses.

Methods

Participants

Data were analyzed from 30 young adults who were recruited from the University of Illinois community and compensated with either course credit or cash. None had participated in Experiment 1. Participants provided written informed consent in accordance with the Institutional Review Board at the University of Illinois. All were right-handed native monolingual English speakers (15 female) with normal or corrected-to-normal vision, no early exposure to a second language, history of brain trauma, or current use of psychoactive medications. Mean age was 19.8 years old (Range: 18 - 22). Mean laterality quotient on the

Edinburgh handedness inventory (Oldfield, 1971) was 0.83 (Range: 0.40-1.00, with 1.0 indicating strongly right-handed and -1.0 strongly left-handed). Fifteen participants reported having a left-handed biological family member.

In order to reach our target of 30 participants, we collected data from a total of 32 participants because two had to be dropped prior to analysis: 1 due to early second language exposure and 1 due to a recording error.

Stimuli

The same stimuli from Experiment 1 were used.

Procedure

The procedure was similar to Experiment 1 except that participants were responding with a button press to categorize answers as correct or incorrect as quickly and accurately as possible. The timing of item presentation was the same, with critical answers being presented only for 200ms. The screen was therefore blank when participants were responding, and the next trial was triggered after a response was registered. Due to the speeded nature of the task, the white "~~~~" break indicators were presented for 2200ms (instead of 1200ms in Experiment 1) because the pacing of the experiment was faster without the forced 1000ms of blank following the answer. As before, the red warning "~~~~" signal appeared for a time within the range of 400 to 600ms.

Electrodes were placed next to each eye and behind each ear in order to record eye movements. This allowed us to remove trials where participants moved their eyes to the lateralized items.

EOG Recording

In order to track eye movements that would invalidate the critical visual field manipulation, electrodes were placed on the face to monitor the electrooculogram. The electrooculogram signal was recorded using 4 silver/silver-chloride electrodes, 2 placed on the outer canthus of each eye, with an electrode on the left mastoid process used as reference and an electrode on the right mastoid process used as ground. All electrode impedances were kept below 5kΩ. The EOG signal was sampled at a 250 Hz and was subjected to an analog bandpass of 0.02 to 100 Hz during online amplification by Sensorium amplifiers.

Raw waveforms were assessed for inclusion on a trial-by-trial basis with artifact thresholds separately calibrated by visual inspection for each subject. Trials were excluded from averaging if they included a horizontal eye gaze movement in the first 200ms of stimulus presentation. Blinks during this time period also were captured because the threshold for trial rejection was set low.

After artifact rejection and including only correct responses, all critical conditions included 27 trials on average, and no individual subject had fewer than 19 critical trials in any given condition.

Behavioral Analysis Methods

Mean levels of accuracy (as raw percentages) will be measured primarily to validate that participants were on task, but they will also be analyzed to check if there are substantial differences across visual fields.

Critical reaction times were trimmed for outliers using a conservative threshold at the 99.75^{th} percentile. Using the function *timefit* from the package *retimes* in R, an ex-Gaussian distribution was then fit to the RT data separately for each subject and condition. The critical distribution parameters (μ, σ, and τ) for each subject and condition were estimated via maximum likelihood estimation through bootstrapped resampling (N = 500) (for review of response time distribution model-fitting, see Van Zandt, 2000). The ex-Gaussian models converged normally for the conditions of interest in the data of each participant.

Separate analyses of μ, σ, and τ for effects of correctness and relatedness across levels of visual presentation mode will be reported.

Results

Accuracy

Table 3.1 reports accuracies for the critical conditions. Across all conditions, accuracy levels were consistently above ninety percent, suggesting that participants were familiar with the arithmetic problems and were appropriately sensitive to accuracy in their responding. The average number of errors committed by a participant across the nine critical conditions was 15, which is about 1.5 errors per condition per participant. For a given condition, only a subset of participants committed errors, and in general the most low-performing participants disproportionately contributed to the results.

Separate analyses were performed to assess if there were any effects of Correctness for two levels of Answer Type (Correct versus Incorrect-Unrelated) or Relatedness for two levels of Answer Type (Incorrect-Related versus Incorrect-Unrelated) with three levels of Visual Field (Left Visual Field/Right Hemisphere, RVF/LH, and Central). There were significant effects of Answer Type for both Correctness ($F(1,29) = 6.12$; $p < 0.05$) and Relatedness ($F(1,29) = 19.61$;

$p < 0.01$), and no effects of Visual Field or interactions between Visual Field and Answer Types, suggesting that accuracies were similar across modes of visual presentation. The effects of accuracy on Correctness and Relatedness were due to higher accuracies overall for Incorrect Unrelated answers (96%, SE = 0.5) compared to both Correct answers (94%, SE = 0.6) and Incorrect Related answers (92%, SE = 0.9).

Reaction Times

Table 3.2 reports the parameter estimates of reaction times for the conditions of interest. Separate analyses of Correctness and Relatedness were performed for each estimated parameter with three levels of Visual Field (Left VF/RH, RVF/LH, and Central). No analyses found any significant effect involving the σ parameter or Visual Field[3].

However, for Correctness, there was a main effect of μ, reflecting that the identification of Incorrect Unrelated answers as incorrect involved longer reaction times (553ms, SE = 17.32) than verifying Correct answers (497ms, SE = 18.50) as correct ($F(1,29) = 4.84$; $p < 0.05$). There was no effect of Correctness on τ ($F(1,29) = 0.26$; $p > 0.05$).

For Relatedness, the reverse was true: there was a main effect on τ, reflecting disproportionate slowing for Incorrect Related answers (462ms, SE = 28.67) relative to Incorrect Unrelated answers (364ms, SE = 22.50) ($F(1,29) = 7.01$; $p < 0.01$), but no main effect on μ ($F(1,29) = 0.05$; $p > 0.05$). These effects are plotted in Figure 3.1, which displays μ and τ for the three answer types collapsed across levels of Visual Field.

[3] Separate analyses of each parameter with only Right Visual Field and Left Visual Field (and not the Central condition) also did not identify any main effects or interactions with viewing condition / hemisphere.

Discussion

In the interest of forming more refined interpretations of Experiment 1, a behavioral experiment was performed wherein new participants responded as quickly as they could without sacrificing accuracy to the critical answer types in each of three visual field presentation modes (left, right, and central). We separately evaluated effects of correctness (comparing responses to correct answers from incorrect unrelated answers) and effects of relatedness (comparing responses to incorrect unrelated answers from incorrect related answers) in these visual fields. Overall, we found high levels of performance (accuracies above ninety percent), and no effects of visual field of presentation on accuracy. There were also no effects of visual field on any measure of reaction time. Importantly, there were several main effects of answer type, suggesting that these VF results were not simply due to the manipulations being ineffective. We will next explain how these findings impact our interpretations of the prior ERP results, first for evaluation of answer correctness and then for answer relatedness.

In the ERP results, there was a P300 that distinguished between correct and incorrect answers at the same latency regardless of visual field. We did not know whether there might be additional processes beyond answer evaluation that differed across the hemispheres and could have resulted in different patterns of RTs across the hemispheres in behavior. Interestingly, there does appear to be evidence for additional response-related processing that is not captured by the P300: the parameter μ (which corresponds to the central tendency of the distribution of reaction times) differed between correct and incorrect unrelated answers, such that incorrect unrelated answers were associated with overall slower reaction times (see Figure 3.1). The magnitude of this effect (around 50ms) would have been readily detectable on P300 latency, so having not obtained a temporally shifted P300 under these circumstances makes the case that something beyond answer evaluation is responsible for the delay. At the same time, the τ parameter (which captures the skew of the distribution towards very slow answers) was *not* significantly different across these conditions. Under some accounts (e.g., Staub & Benatar, 2014; Reingold, Sheridan, & Reichle, 2015; Jackson, Balota, Ducheck, & Head, 2012), shifts in τ reflect additional processing and/or disruptions in processing. Thus, our result pattern shows a modal shift in RT to incorrect, relative to correct, answers, which seems to arise after stimulus evaluation, and might reflect influences from criterion-setting that apply uniformly, as opposed to a distributional shift due to additional processes coming online for a subset of trials.

The μ effect obtained for correctness did not interact with presentation condition (i.e., it was the same across the hemispheres). Thus, not only do readers use the same amount of time to evaluate each provided answer regardless of which hemisphere was preferentially processing it, we also found no evidence that there are additional differences in response selection and execution to verify answer correctness as a function of visual field in this paradigm. These results are in contrast to reports from brain damaged participants and some split-brain patients (Gazzaniga & Smylie, 1984; Funnell, Colvin, & Gazzaniga, 2007; Delazer, et al., 1999), which have found that the arithmetic abilities of the isolated LVF/RH are poor, but are consistent with studies using larger groups of non-brain damaged participants (Andres et al., 2011), which have suggested that the RH does contribute to normal processing of arithmetic problems.

The other effect of interest in the present study was the sensitivity of the distributional parameters to the manipulation of answer relatedness. Relatedness effects are typically reported as slower mean RTs to incorrect answers that are related to the correct answer compared to unrelated incorrect answers (see Ashcraft, 1992, for review). We found that this effect manifested on the parameter τ in the form of a longer τ in the incorrect related condition as compared to the incorrect unrelated condition (see Figure 3.1). This finding is entirely consistent with what would be expected given the current literature on RT distributions: in more cognitively demanding or working memory-intensive situations, effects of τ are frequently obtained rather than effects of μ (e.g., Balota & Spieler, 1999; McVay & Kane, 2012; Payne & Stine-Morrow, 2014). Thus, by examining the underlying RT distribution rather than simply reporting the mean, we found additional evidence in support of claims that rejecting incorrect related answers involves the recruitment and utilization of additional processing mechanisms (e.g., inhibitory control and working memory) beyond what are needed to reject incorrect unrelated answers (Barrouillet et al., 1997; Rammelaere et al., 1999; Lemaire et al., 1996).

Interestingly, there was no difference in μ between incorrect related and incorrect unrelated answers, creating an apparent one-to-one mapping between effects of correctness and relatedness, and the distributional properties of μ and τ, respectively. That is, we identified evidence differentiating between effects of answer correctness (μ) and effects of answer relatedness (τ), suggesting that the processes involved in their generation might not be entirely overlapping -- i.e., we show a new, qualitative difference in the nature of the effects that was not evident from prior work showing just that RTs to these conditions were longer on average. This is also consistent with what ERPs had revealed about arithmetic processing; there, the effect of correctness was present on the P300, whereas the effect of relatedness only emerged on a late positivity.

We were also interested in how relatedness effects might manifest differently in behavior across the hemispheres, given that the late positivity difference for relatedness in the ERPs was only significant in the LVF/RH. That ERP result suggested that the RH performed some additional cognitive processing of the answers, which, given its nature and timing, we interpreted as deliberative in nature (rather than a more automatic response). However, in the behavioral results, the effect of relatedness (on τ) did not differ as a function of presentation condition or VF in particular (just as μ, which was sensitive to answer correctness, was similar across visual fields).

Given that the only evidence of sensitivity to relatedness that we had in our brain measure was following LVF/RH presentation, the main effect of relatedness in behavior suggests that the ERP relatedness effect was non-trivial (i.e., is functionally operative). That is, although only the RH showed evidence of relatedness effects in ERPs, in behavior, both the left hemisphere and the joint processing of both hemispheres (central presentation) also showed appreciation of this more detailed aspect of the provided answer. Mechanistically, whatever cognitive process led to the generation of the late positivity might have interactively spread from the right hemisphere, ultimately impacting decision making under all viewing conditions. This is consistent with many theories of interhemispheric communication that claim that the RH rapidly shares information with the LH (e.g., Marzi, Bisiacchi, & Nicoletti, 1991; Barnett & Corballis, 2005). However, it is also possible that interhemispheric transfer was required in order to perform the response task, as manual responses necessarily require access to relevant information to execute the response, and, across trials, both hands were used (mapping of the yes/no response type for R/ L hand was counterbalanced across participants).

Regardless of the source of the interhemispheric transfer, our results suggest that the decision-making information that originated from the LVF/RH (as suggested by the ERP results) spread to the LH and affected overall behavior. That the process indexed by the late positivity could conceivably *cause* the behavioral effect is at least consistent with a comparison of the latency of the late positivity, which peaked between 400 and 600ms post-stimulus onset, and the average raw RTs observed for the relatedness condition, which occurred at 1045ms. In future work, it would be useful to assess this possibility at a trial level with concurrently recorded behavioral and ERP results.

Beyond the opportunity to expand our understanding of the mechanisms involved in answer evaluation by uncovering novel distributional features of the response times obtained in this task, perhaps the biggest takeaway from the present work is that there were no LH advantages in answer evaluation and response execution at all. This is very different from what is typically reported in behavioral language studies, wherein the LH is generally faster/more accurate (e.g., Jordan et al., 2000). Thus, in healthy young adults, we find that basic answer evaluation and decision making is not the sole domain of LH processing (as has been suggested by, e.g., Dehaene et al., 1999; Dehaene & Cohen, 1995), but, instead, that the RH is actively involved in evaluating and executing responses for these answers as well. Interestingly, the RH has previously been associated with the ability approximate answers and perform basic number comparison (e.g., judging relative quantities for size) – both kinds of skills which may be possible to perform through more deliberative evaluation. Consistent with this view, in patients with LH brain damage (who thus must rely on RH processing), it is often reported that they employ slow, elaborative or procedural techniques to reach the correct answer (Crutch & Warrington, 2002; Diesfeldt, 1993; Cappelletti, Butterworth, & Kopelman, 2001, 2012). Taken together, it is possible the behavioral relatedness effect we are observing here is a manifestation of that same particular feature of RH processing that is observed in patients.

Figures and Tables

Table 3.1: Accuracy

	Accuracy					
	Left VF Right Hemisphere		Right VF Left Hemisphere		Central Viewing	
	Accuracy	*SE*	*Accuracy*	*SE*	*Accuracy*	*SE*
Correct	92.99	1.32	94.81	0.85	95.69	0.92
Incorrect – Unrelated	95.80	0.98	96.67	1.05	96.92	0.58
Incorrect – Related	93.23	1.19	92.40	1.48	90.47	1.74

Table 3.1: Participant accuracy at categorizing an answer as correct or incorrect across viewing conditions.

Table 3.2: Reaction Time Parameter Estimates

		Reaction Time Parameter Estimates					
		mu (ms)	SE	**tau (ms)**	SE	**sigma (ms)**	SE
Right VF (LH)	Correct	502	31	399	43	76	15
	Incorrect-Unrelated	539	28	376	41	67	13
	Incorrect-Related	536	37	486	54	95	40
Left VF (RH)	Correct	526	36	387	43	90	22
	Incorrect-Unrelated	553	33	369	40	61	9
	Incorrect-Related	550	27	470	52	131	50
Central Vision	Correct	463	29	358	47	68	11
	Incorrect-Unrelated	566	29	348	36	56	11
	Incorrect-Related	555	35	430	44	82	22

Table 3.2: Estimates of mu, tau, and sigma across viewing conditions and stimulus category types.

Figure 3.1

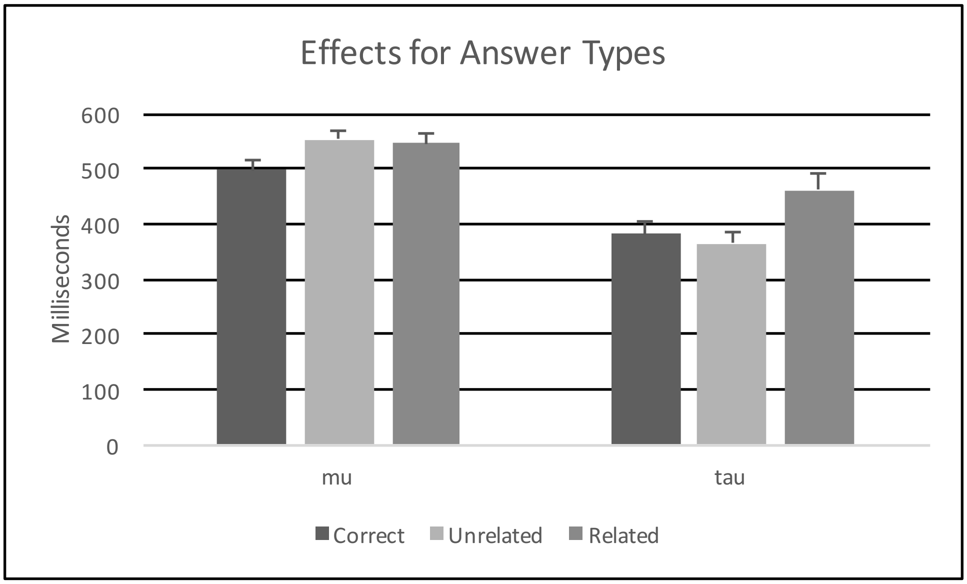

Figure 3.1 Estimates of mu and tau across answer types. The effect of correctness can be seen in mu, whereas the effect of relatedness can be seen in tau.

CHAPTER 4

Experiment 3: On the potential (in)flexibility of semantic memory access for numerals: the effect of incidental repetition across tasks

Introduction

In Experiment 1, we observed that effects of answer expectancy occurred relatively early following stimulus presentation – just after 300ms. Both the timing and the morphology of the expectancy effect in arithmetic expressions was inconsistent with what is typically observed in sentence processing paradigms, wherein there is a larger negativity for unexpected words relative to expected words that peaks just prior to 400ms post-stimulus – i.e., an amplitude modulation of the N400, a component linked to amodal semantic access. This difference in processing mechanisms across domains is not intuitive because both sentences and arithmetic expressions seemingly construct a meaningful message about what is likely to appear next – yet this meaning seems to not be appreciated the same way in arithmetic expressions as it is in not only language but also other, semantically richer domains (as N400 congruity effects obtain not just for words following sentential contexts, but also, e.g., in paradigms with unexpected pictures and movie clips: Nigam, Hoffman & Simons, 1992; Sitnikova, Kuperberg, & Holcomb, 2003).

One possible explanation for the apparent absence of N400 effects in arithmetic contexts is that component overlap – in particular, the co-occurrence of the P300 – creates a barrier to directly measuring the N400. The P300 is particularly difficult to disentangle from any potential effect on the N400 because the direction of the effect is the same: more positive (less negative) for target (expected) items than non-target (unexpected) items. Indeed, although there are many reports of N400 effects in arithmetic contexts, these almost universally occur in situations where there are positivities in response to the correct answer. Even in Experiment 1, in which the P300 effect preceded the typical timewindow for the N400, it is difficult to assess what might be happening in the N400 timewindow in the absence of the prior effect (i.e., the P300 peaks around the time that a divergence on the N400 would be expected to start).

Due to the overlap with the P300, it is not clear at present how thoroughly numerals (specifically, the Arabic system of representing number values) are processed for their conceptual meaning. The N400 is thought to reflect a first pass into long-term semantic memory, and larger amplitude N400s are associated with more new processing of a semantic concept (Kutas & Federmeier, 2011). What is typically examined is the amount of facilitation (reduction in N400 negativity) an item receives as a function of prior semantic processing, similar to conceptual priming. When it was first proposed that arithmetic answers were processed for their semantic content similarly to words in sentences, the implication was that our memory for Arabic numerals is structured and accessed similarly to how the conceptual relationships between words are processed, at least when probed in the context of arithmetic expressions (Niedeggen & Rösler, 1999; Niedeggen, Rösler, & Jost, 1999). However, given the difficulty in isolating the N400 component in these studies, it is unclear at present whether that claim is fully justified.

A critical question to resolve, then, is if Arabic numerals convey the type and degree of conceptual information needed to influence N400 amplitude. One possibility is that these symbolic representations of numerical information are not processed for meaning the way that words and pictures seem to be, and that effects of contextual support from numerals thus do not result in modulations of the N400 on subsequent numerals. It is already known that not all forms of expectancy in familiar, and potentially meaningful, sequences are associated with N400 effects. For example, there is a long history of searching for N400 effects in response to deviant notes in familiar musical melodies – and failing to find them (Besson & Macar, 1987; Paller, McCarthy, & Wood, 1992). It is possible that the conceptual information conveyed by arithmetic expressions might be more similar to the case of learned musical patterns/sequences than to the case of sentences. That is, arithmetic expressions might convey too impoverished a semantic message to elicit the kind of N400 effects that are seen for sentences, picture sequences, and movie clips. If true, it would be consistent with the interpretation that, indeed, effects of expectancy in purely arithmetic contexts (using Arabic numerals rather than number words) are not modulations of the N400, but are instead limited to differences in P300 amplitude.

In order to assess whether double-digit Arabic numerals of the sort usually found in common arithmetic problems can elicit N400 effects (and thus access semantics similarly to other stimulus types), we will utilize a paradigm in which N400 effects can be generated in the absence of an overlapping P300. Specifically, a simple design that allows for the examination of N400 effects in the absence of componentry related to evaluation and decision-making (i.e., P300) is a repetition paradigm in which participants are unable to render decisions when presented with the critical (repeating) numerals and, instead, decision-related information is presented in a delayed probe item. Through this method, rather than measuring the amount of facilitation an item receives as a function of a prior arithmetic expression, we will instead be simply measuring the amount of facilitation an item receives as a function of its own prior presentation earlier in a list (i.e., a self-priming effect). In sum, this approach could reveal evidence of N400 effects to common Arabic numerals, suggesting that indeed there is some obligatory meaning being accessed when we view these numerals.

Importantly, we will not be measuring facilitation to immediate and direct repetitions (which could reflect perceptual priming in addition to any semantic priming that might also occur), but on incidental, later repetitions, such that the conceptual meaning of the numeral must be retained in memory through several intervening trials. This is a widely used method for examining the quality of semantic processing a potentially meaningful item undergoes, and reductions in N400 amplitude in these contexts are more broadly thought to reflect conceptual priming of items (e.g., Voss & Paller, 2006; Voss, Schendan & Paller, 2010; Voss & Federmeier, 2011). That such delayed repetition effects reflect conceptual priming is supported by evidence that deliberate manipulations of the amount of conceptual processing influences the size of the effect. For example, although N400 priming effects are often small or absent for strings of meaningless letters (e.g., “TXQ”), if participants are encouraged to entertain the possibility that these items are meaningful by asking them to judge whether or not they are a familiar acronym, then illegal strings manifest repetition effects similar to other, meaningful strings of letters (e.g., words and acronyms; Laszo, Stites, & Federmeier, 2011). Similarly, although N400 effects are difficult to find when people are processing melodies, it is possible to obtain them through cross-modal priming studies with language stimuli and other sophisticated manipulations that tap into richer aspects of the meaning of musical elements (Daltrozzo & Schön, 2009; Koelsch et al., 2004). As such, even if it is the case that arithmetic expressions do not convey sufficiently robust semantic messages to elicit N400 congruency effects, then some other task that encourages access to different, broader associations with numerical information might be able to draw out their conceptual meaningfulness.

Motivated by this possibility, our examination of the processing of symbolic representations of number (i.e., Arabic numerals) for conceptual information will include three tasks, each of which recruit different types and amounts of engagement with these numerals' potential semantic content. The effects of repetition will be examined in order assess how much meaning is extracted from numerals across these different task contexts. Critically, what is measured across these tasks is always the response to first and second presentation of numerals (not responses to the subsequent probe items, which vary across task). This enables us to directly measure, within subjects, how the size and morphology of the repetition effect to the exact same physical stimuli is influenced by the kind of processing demanded by each task. The three tasks and their respective potential for eliciting repetition effects for Arabic numerals are described below.

In the "matching task", relatively little assessment of the conceptual content of the numeral is required. Participants view a double-digit number and later verify the presence or absence of a provided probe digit – essentially, performing a delayed detection (or match-to-sample) task. This encourages participants to see numerals as objects with subparts, but does not require them to access their memories of those objects for additional conceptual information about them. The critical question this task asks is whether or not numerals are obligatorily processed for their conceptual meaning. In similar paradigms using words, N400 repetition effects have been observed, showing that meaning extraction is relatively automatic (although this N400 repetition effect is smaller than when reading for semantic category membership; Rugg, Furda, & Lorist, 1988). However, it has also been observed that for other items that are less inherently meaningful than words (e.g., ambiguous and novel line drawings), simply detecting visual features does not result in N400 effects, and, instead, such effects are only obtained when participants are encouraged to think about these stimuli as potentially meaningful (Voss et al., 2010). Arabic numerals are presumably more meaningful than shapes that viewers have not previously experienced before, but whether the type of meaning elicited by numerals is more similar to wordforms (and meaningful pictures), which typically elicit robust N400 effects in repetition paradigms, versus novel shapes or scrambled images, which typically do not elicit robust N400 effects, is unknown. In sum, if access to semantic knowledge of numerals is as obligatory as it seems to be for words, then this task should result in N400 repetition effects (albeit perhaps smaller than a more conceptually driven task).

In the "divisor task", participants view a numeral and then must later verify whether a probe number is one of its factors (e.g., seeing "14", participants would confirm probe items of "1", "14" "2", and "7"). This task potentially requires participants to access more of their knowledge of numbers than the first task. If retrieving associated arithmetic facts activates a semantic network for the numeral itself, or otherwise conceptually primes the presented numeral, then one would expect facilitation on the N400. This task thus most directly addresses the question of whether or not the expectancy/congruency effects observed in arithmetic contexts reflect modulations of N400 amplitude – but critically, since the actual arithmetic probe comes later in this paradigm, we can look at the response to the numeral without concurrent P300 activity.

Finally, in the "quantifier task", we attempt to emulate typical "read for comprehension" tasks often used in paradigms studying the semantic processing of words. It is unclear with numerals what the true equivalent of this might be, since the significance of Arabic numerals as a representation can vary as a function of use (e.g., as part of an arithmetic problem, as a label for the identity of a sports figure, as a representation of associated magnitude). However, under the assumption that most people's natural experience with numbers outside the realm of the classroom is in the context of measuring and counting, we had people read numerical expressions (e.g., "85" followed by the quantifier "yards") and judge them for plausibility. This task is aimed at tapping into their conceptual knowledge of the size/magnitude conveyed by the numeral, and presumably also engages other long-term memories of experiences with it. We expected that, if any task would be able to demonstrate conceptual priming for Arabic numerals in the form of N400 facilitation, this would be the one most likely to succeed.

Stepping back from the particular nuances of each task and their respective possibility for eliciting effects of repetition, the main question being addressed by this approach is whether Arabic numerals are able to convey a rich enough semantic concept to modulate the N400, a component functionally linked to processing the meaningfulness of a stimulus. In general, the approach of using single-item presentations with repetitions of identical or related primes has been critical for honing our understanding of the temporal dynamics of extracting various properties (such as semantics) from a diverse group of stimuli across modalities, such as visual and auditory words (e.g., Rugg, Doyle, & Melan, 1993; Joyce, Paller, Schwartz, & Kutas, 1999; Holcomb & Grainer, 2006; 2007), faces (e.g., Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002; Schweinberger, Huddy, & Burton, 2004), dynamic gestures (Corina, Grosvald, & Lachaud, 2011; Wu & Coulson, 2005), environmental sounds (Shirmer et al., 2011), and pictures of objects (McPherson & Holcomb, 1999; Schendan & Kutas, 2003; 2007; Eddy, Schmid, & Holcomb, 2006), so this study will support and extend the existing broader literature of how different classes of stimuli are assessed and processed for meaning.

Methods

Participants

Twenty-four participants were included for analysis in the study (due to the relatively large numbers of trials per critical condition, stable effects should not require more than this). All were right-handed native monolingual English speakers (12 female) with normal or corrected-to-normal vision, no early exposure to a second language, history of brain trauma, or current use of psychoactive medications. Participants' average age was 20 years (range: 18 – 25), their mean laterality quotient on the Edinburgh handedness inventory (Oldfield, 1971) was 0.79 (range: 0.39 - 1.00, with 1.0 indicating strongly right-handed and -1.0 strongly left-handed), and 10 reported having a left-handed biological family member. In order to reach the number of 24 participants, data was additionally collected from 2 participants who were dropped prior to analysis due to failure or inability to comply with instructions (i.e., excessive sleep deprivation and/or blinking during critical trials). Informed consent was received from all participants in accordance with the Institutional Review Board at the University of Illinois.

Stimuli

Critical Stimuli for All Tasks

All double-digit numbers were used as critical stimuli, with fillers being either single- or triple- digit initial numbers (1-9; 100-116). From the perspective of the participant, numbers that they were attending and responding to ranged from 1 to 116. Critical numbers repeated after 2 or 3 intervening numbers. Filler items did not repeat; this was necessary to build the lag structure. There was no indication that numbers in the double-digits were of special interest other than that they repeated after some intervening items. To avoid unintentional repetitions of non-identical numbers, there were no consecutive numbers with matching initial digits (e.g., 90 would not be directly followed by 93), and sequential matches of the second digit were rare.

Three separate orders of numbers and structures of repetitions were generated, so numbers that had 2 intervening items in one list might have 3 intervening items in another list, and numbers which appeared chronologically earlier on one list might appear later on another. This was necessary to have all subjects be able to perform all three tasks, and it also provided experimental security so that particular nuances of a given list structure could not be responsible for any effects.

Probe Stimuli for Matching and Divisor Tasks

In the Matching and Divisor tasks, critical numbers were followed by a single-digit probe. To prioritize having the correct response be evenly distributed between YES/NO, distribution of these probes differed across these experiments (i.e., there were a disproportionate number of probes of the digit 2 in the Divisor task). In the Matching task, the probe query required attention to have been paid to any one of the digits in the critical item in an unpredictable manner. Probes for correct answers were distributed as evenly as reasonably possible between the first digit and second digit.

Probe Stimuli for the Quantifier Task

In the Quantifier task, the critical numbers were followed by a probe quantifier word. These words included basic units of measurement (e.g., inches, gigabytes), as well as some collective nouns (used in an incomplete expression fragment, e.g., “pieces of…”), arrangement classifiers (“rows of…”, “pieces of…”) and varietal classifiers (“stages of…”, “types of…”). The use of expression fragments was necessary to encourage the measurement/counting interpretation of the word rather than its basic noun meaning.

In order to select quantifier words that were best associated with measurements and numbers, 105 English-speaking participants from the United States were recruited to participate in a norming study through Mechanical Turk. These participants consented and were compensated with $1.50 in accordance with University of Illinois IRB policies. In the norming study, participants were given a sliding ruler that spanned 0 to 100, and their assignment was to select a number that felt "most natural" completing the expression or word they were provided with. For example, they saw, "______ calories" and selected the number that they most naturally fit with the given measurement term. Critically, there was also a "No number feels natural" option, which was used to determine whether or not a given term would be sensible to our participants in the ERP experiment.

Data from 101 of 105 subjects was usable; a few participants either skipped too many responses or picked "No number feels natural" for more than half the items. There were 100 potential words for measurement tested for viability. The threshold for exclusion was if five or more participants reported "No number feels natural" for a given word. This process removed 53 words from the set, leaving 47 to be used in the study.

Of the 47 quantifiers included, there was variability in the numbers that were selected to be natural-sounding with them. The smallest average response was for "teaspoons" (mean: 6.30, mode: 2.0) and the largest average response was "calories" (mean: 87.95, mode: 100.0). The average of all responses was around 30 (29.19), and the majority of responses were below 50. Respondents only picked numbers above 50 for the following terms: gigabytes, pounds, megabytes, degrees Fahrenheit, and calories. The default position of the sliding ruler was placed at 50 (and to be registered as a response, they would have to move it deliberately off of this position), so this bias towards being comfortable counting smaller numbers was not because the slider would need to be dragged farther to reach a higher number.

On all lists, probe words repeated 4-5 times (average: 4.4), and there was a buffer of at least 15 intervening items before a given probe word would be used again. This was necessary because there were 206 items to be responded to and only 47 probe words that met criteria. As will become important later, this feature also allows ERP repetition effects for normal words to be measured in a within-subject design.

Procedure

General

In all tasks, the timing of items followed the same format. Prior to the critical stimulus, a series of red "~~~~" were presented to alert the participant that a number was coming up. To discourage complete predictability of stimulus presentation, this remained on the screen for a random period of time within the range of 400 and 600ms. The critical stimulus was on the screen 500ms, followed by 800ms of blank to allow for ERP data collection. The probe query stimulus was then displayed for 500ms, after which the screen went blank. Subjects could respond as soon as they were able; speed was not emphasized. After a response was registered, a series of white "~~~~" were presented for 1500ms during which subjects could blink and move their eyes around. The red "~~~~" directly followed this blinking period. A small square was continuously presented in the center of the screen just below the stimuli to remind participants where to keep their eyes during otherwise blank periods.

There were periodic self-paced breaks in which yellow "~~~~" were presented after a response – as in the prior Experiments, this indicated that the subjects could take a break until they were ready to do more trials. During each of the three tasks across the experiment, there were also two experimenter-controlled breaks: one after an initial small batch of trials to check in on the subjects' understanding of that task, and one roughly halfway through. Each repetition structure list had slightly different structure of breaks because interruptions are not possible when repetitions are in progress.

Procedures that were unique to each task follow.

Matching Task

After reading the critical initial number, participants were probed with a second (single-digit) number. They responded YES/NO to this number depending on whether that probe number was one of the digits in the first number. On trials in which the initial (filler) number is a single digit, participants were just answering whether the number repeated. More often, the first number was a double- or triple-digit number whose parts or identity needed to be held in memory.

Divisor Task

After reading the critical initial number, participants were probed with a second (single-digit) number. They responded YES/NO to this number depending on whether the first number was divisible by the second number (i.e., if the probe was a factor). Participants were specifically instructed to assess if they could "take the first number, divide it by the second number, and get a whole number". They were further instructed that they did not need to know what the answer was, only whether or not it was a whole number.

Quantifier Task

After reading the critical initial number, participants were probed with a quantifier word or expression that is typically used for measuring or counting items (as indicated by the prior norming study). Their instructions were to judge whether or not the previous number "seems typical for a given type of measurement" and to "decide if you think this is a common or typical pairing." They were given examples of both rarer and more normal pairings, and it was further explained that there were no right answers because their judgments would depend on their own experience and familiarity with how numbers are used to express distances, sizes, amounts, and other measurements. All subjects read the same instruction sheet, and very little additional instruction was necessary despite the novelty of the task. After presentation of the word, subjects responded either "reasonably common" or "reasonably uncommon".

EEG Recording

Brainwaves were recorded using 26 silver/silver-chloride electrodes mounted in an elastic cap. All electrode impedances were kept below 5kΩ, and electrodes were evenly distributed over the scalp (see Figure 2.1 from Experiment 1 for the configuration). The data were referenced on-line to the left mastoid and re-referenced offline to an average of the right and left mastoids. A separate frontal electrode acted as ground. The vertical electrooculogram (EOG) signal was monitored with an electrode just below the eye (on the infraorbital ridge) and horizontal EOG was monitored with electrodes placed on the outer canthus of each eye. The EEG signal was sampled at a 250 Hz and was subjected to an analog bandpass of 0.02 to 100 Hz during online amplification by Sensorium amplifiers.

Raw waveforms were assessed for inclusion on a trial-by-trial basis with artifact thresholds separately calibrated by visual inspection for each subject. Trials were excluded from averaging if they included blinks, movement artifacts, signal drift, blocking, or a horizontal eye gaze movement. After artifact rejection, critical bins included 82 trials on average (range: 55 - 90).

Epochs of EEG data for each trial were taken from 100ms prior to item onset until 920ms after item onset, and the baseline acquired over the 100ms preceding the onset of each trial was subtracted prior to averaging. ERP mean amplitudes were measured after application of a digital bandpass filter of 0.1 to 20 Hz. In all of the results below, interactions and effects of electrode-related factors are only reported for those of theoretical relevance.

Results

General behavior

Accuracy for responding to probes was high in both the Matching Task (98.8%) and Divisor Task (93.4%). In the Quantifier Task, the average percent of numerical expression endorsement was 54.7%, indicating a slight bias to respond "Yes" (lowest endorsement rate: 22.3%; highest endorsement rate: 81%).

Event-related potentials

Initial presentation across tasks

Prior to explicating the effects of primary interest to the current report (i.e., repetition effects on the N400), we first examined the general effect of task on the processing of the critical (non-probe) numbers. Visual examination of the waveforms (Figure 4.1) suggested a sustained effect differentiating the Quantifier Task from the other two tasks. To examine the pattern, we collapsed across the factor of repetition and compared the amplitude of the response to numerals in the Matching, Divisor, and Quantifier Tasks directly. We computed an average over a broad window, from 300-700ms post-stimulus, and performed an ANOVA with 16 channels to cover the following distributional factors: 2 levels of Hemisphere (Left Hemisphere, Right Hemisphere), 2 levels of Laterality (Lateral, Midline), and 4 levels of Anteriority (from front to back). There was a main effect of Task ($F(2,46) = 4.39$, $p > 0.05$), which did not interact with any distributional factors, consistent with the widespread appearance of the effect.

An analysis over all channels found that this effect of Task was due to more negativity in the Quantifier Task (1.19μV, SE = 0.07) than during either the Matching (1.66μV, SE = 0.07) or Divisor (1.82μV, SE = 0.08) Tasks. Post-hoc pairwise comparisons over all channels found significant differences between the Quantifier Task and both the Matching Task ($F(1,23) = 5.70$, $p < 0.05$) and the Divisor task ($F(1,23) = 14.58$, $p < 0.01$), but no significant difference between the Matching and Divisor Tasks themselves ($F(1,23) = 0.61$, $p > 0.1$).

Repetition effects in the N400 timewindow

An analysis of the latency of the N400 across tasks (in a timewindow from 250ms to 500ms post-stimulus) found its peak on average to be at 395ms (SE = 1.2ms). An omnibus ANOVA was therefore performed on 16 channels around the center of the N400 effect, from 300 to 500ms post-onset of critical number, with 3 levels of Task (Matching, Divisor, and Quantifier), 2 levels of Repetition (First Presentation, Second Presentation), and, to characterize the scalp distribution of any effects, 2 levels of Hemisphere (Left Hemisphere, Right Hemisphere), 2 levels of Laterality (Lateral, Midline), and 4 levels of Anteriority (from front to back).

This analysis revealed significant main effects of Task ($F(2,46) = 5.46$, $p < 0.01$) and Repetition ($F(1,23) = 5.40$, $p < 0.05$), and an interaction between Repetition, Hemisphere, Laterality, and Anteriority ($F(3,69) = 3.83$, $p < 0.05$). (Other interactions, between Task and Hemisphere ($F(2,46) = 5.77$, $p < 0.01$), Task and Laterality ($F(2,46) = 3.24$, $p < 0.05$), Task, Hemisphere and Laterality ($F(2,46)=3.26$), $p < 0.05$), and Task, Laterality, and Anteriority ($F(6,138) = 2.58$, $p < 0.05$), are reflective of the longer-lasting main effect previously characterized).

Main effect of repetition

The main effect of repetition was due to facilitation (i.e., less negativity) for Second Presentation (repeated items) (1.72μV, SE = 0.07) relative to First Presentation (unrepeated items) (1.47μV, SE = 0.07). This effect's interaction with distributional factors was due to relatively more facilitation for repeated items in centrofrontal channels (somewhat left-lateralized) relative to very posterior and lateral posterior channels in general; see Figure 4.2 for a visual representation of this effect and its topographical distribution.

Effects of repetition in each task

Planned comparisons to examine the robustness of the effect of repetition in each task were performed over 21 channels (dropping the very posterior channels where there was no main effect). In the Matching and Divisor Tasks, there was no significant effect of repetition ($F(1,23) = 2.43$ and 2.05 respectively, p's > 0.1). This was also true over a smaller, more focal selection of channels. However, as expected given that there was no interaction with task, there was a trend towards a repetition effect in both tasks. In the Matching Task, Second Presentation (1.49μV, SE = 0.10) was numerically more positive than First Presentation (1.20μV, SE = 0.09), and the same was true of Second Presentation (1.70μV, SE = 0.10) relative to First Presentation

(1.42μV, SE = 0.10) in the Divisor Task. Critically, in the Quantifier Task, there was a significant effect of repetition ($F(1,23) = 5.13$, $p < 0.05$), due to more facilitation for Second Presentation (1.12μV, SE = 0.08) compared to First Presentation (0.80μV, SE = 0.08). Figure 4.3 displays the results of this analysis in the form of a series of bar graphs, which collectively show that there is a remarkable similarity in the (small) size of the repetition effect across all tasks.

Effects of repetition for quantifier words

To mitigate against potential concerns that the participants in this study were unusual in some way (that then resulted in a very small repetition effect with an unusual distribution), we took advantage of a feature of the Quantifier Task: the probe items were normal words, and all of the words repeated (i.e., were re-used later). Thus, we have access to a within subject measurement of word repetition effects in the present sample. There were at least 15 intervening trials prior to a word being repeated (substantially more than the number of intervening items between critical numerals). After artifact rejection, each presentation of the critical word had on average 34.5 trials per subject.

We performed an omnibus ANOVA over 16 channels around the center of the N400, from 300 to 500ms post-onset of the critical probe word, with 2 levels of Repetition (First Presentation, Second Presentation), and, to characterize the scalp distribution of the expected effect, 2 levels of Hemisphere (Left Hemisphere, Right Hemisphere), 2 levels of Laterality (Lateral, Midline), and 4 levels of Anteriority (from front to back). This analysis found no main effect of Repetition ($F(1,23) = 1.31$, $p > 0.05$), but did identify an interaction between Repetition and Anteriority ($F(3,69) = 13.75$, $p < 0.001$).

Follow-up analysis was performed within two levels of Anteriority, frontal (with 11 electrodes) and posterior (with the other 15 more central and posterior channels, where word repetition effects typically are obtained). There was no effect of Repetition over frontal electrodes ($F(1,23) = 0.46$, $p > 0.05$). However, over more posterior electrodes there was an effect of Repetition ($F(1,23) = 9.45$, $p < 0.01$), with words at Second Presentation (3.12 μV, SE = 0.22) being more facilitated (positive) than words at First Presentation (1.88 μV, SE = 0.21). A representative channel displaying this effect is plotted in Figure 4.4, along with a topographical display of its distribution. In Figure 4.5, the size of this repetition effect (i.e., the difference between first and second presentation) is plotted in direct comparison with the same effect for numerals.

Discussion

The purpose of the present set of experiments was to clarify whether or not Arabic numerals convey a stable enough semantic representation to influence semantic access upon repetition. In general, we found support for this possibility: reliable N400-like effects were elicited by the repetition of these numerals. In particular, we found that a negative deflection in the waveform, which peaked right before 400 ms, was reduced in amplitude (more positive) upon repetition. Under the assumption that this amplitude reduction we observed during the traditional N400 timewindow represents semantic access, we thus found evidence that there is less new semantic activity for a repeated numeral compared to the semantic activity present during its initial presentation.

Our second question was whether these effects would be influenced by the level or type of semantic access that was being performed on the numerals: would a more superficial level of processing elicit the same kind of effect as a more conceptually driven level of processing? Despite using a range of tasks that seemed to require different levels and types of information access, there was no interaction between the effect of repetition and task, because all tasks elicited numerically similar effects (see Figure 4.3). Although we cannot rule out subtle differences in the size or distribution of the effects elicited in the different tasks, the observed pattern suggests that the amount of meaning gleaned from processing a number seems remarkably stable. If this repetition effect does belong in the family of N400s, it would suggest a high level of stability in the amount/type of semantics that is accessed from symbolic numerals.

There are at least two features of this effect of repetition that seem different from what is typically observed for repetition of other kinds of meaningful stimuli. One is that it is a very small effect – only about 0.30µV separates first and second presentations. In studies using words and letter strings, the difference that is reported tends to be substantially larger, around 1µV or 1.5 µV (see, e.g., Laszlo & Federmeier, 2007). Further, the interitem lag (i.e., number of intervening items, which here was 2 or 3) is also on the smaller end than what has been used to observe repetition effects for words, which, if anything, should result in larger effects of repetition. In general, it is the case that, in contrast to words, the extent to which Arabic numerals receive a benefit from repetition seems smaller. Indeed, probe words that repeated in the quantifiers task had a more substantial repetition effect despite having notably longer repetition lags (see Figures 4.4 and 4.5).

Given the parameters of the present study (high trial numbers, low numbers of intervening items), that the repetition effects were small indicates that numeral repetitions are associated with only a small amount of conceptual priming. Prior effects reported for expectancy-based "semantic" facilitation in arithmetic contexts have consistently been substantially larger than 0.3 μV (e.g., in Exp. 1 were around 1 μV in central presentation, which is on the small side for that domain), and are also distributed more posteriorly than the present N400-like repetition effect. Thus, it would seem that the vast majority of those effects are, at least primarily, due to modulations in P300 amplitudes, although it is possible that there could be a very small contribution from N400 reductions as well. Certainly, the present results are inconsistent with prior researchers' interpretation of the congruity effects in arithmetic contexts as entirely N400-based, given how relatively small numerals' sensitivity to contextual support is – even from prior presentations of the same representation.

The second major distinguishing feature of the repetition effect we obtained for numerals is its scalp distribution. In word repetition paradigms, the distribution of the effect tends to be centroparietal (typical for an N400 effect using an average mastoid reference) – indeed, the effect of repetition of words in these participants also had this canonical scalp distribution (see Figure 4.4). However, in the case of Arabic numerals, we found that the repetition effect distribution is more frontal and left-lateralized (see Figure 4.2). Although the effect was temporally and functionally similar to an N400, this atypical scalp distribution raises the question of whether the obtained repetition effect should be classified as an N400.

It has been reported that N400 effects can shift distribution across stimulus categories, although precise characterization of these shifts has been difficult because of component overlap that often exists in paradigms eliciting N400 responses to other (non-word) stimulus types. For example, although N400 effects involving pictures and line drawings are typically more frontally distributed than those to words, pictures also tend to elicit a frontally distributed N300 effect (associated with the processing of visual structure), as well (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999), which could affect the measured N400 distribution. Other studies, however, have found evidence for a distributional shift in the N400 to pictures that does not seem to be attributable to N300 overlap (e.g., Ganis, Kutas, & Sereno, 1996; Ganis & Kutas, 2003). Here, we similarly have a distributional shift that would not appear to be due to component overlap (especially since letter strings and digit strings have highly overlapping perceptual properties). As this kind of distributional shift is generally taken to indicate that at least partially non-overlapping sets of neural generators were involved in the responses, this pattern thus provides indirect evidence that the nature of the semantic representations being accessed is different for numerals and for words (to the extent that different sources imply different kinds of semantics).

We also observed an unexpected main effect of task in the form of a long-lasting negativity in response to numerals when they were about to be followed by a quantifier, compared to the tasks with digit probes (see Figure 4.1). This effect could be part of a larger class of negativities that have been characterized as being sensitive to uncertainty about upcoming stimuli. The quantifier task is associated with relatively more uncertainty about how to respond to the upcoming probe, as there are no clear yes/no answers, and as participants, rather than expecting one of 9 possible single digit probes, could not form much of an expectation for what specific probe words they might see. Thus, this main difference across tasks in response to numerals might not have to do with processing of the numerals themselves, but rather with the requirement to wait for an upcoming probe (see Brunia, van Boxtel, & Böcker, 2012, for review of anticipatory negativities).

However, effects of this nature typically last the entire retention period, whereas in the present data the negativity resolves after a few hundred milliseconds. Alternatively, then, this effect might just be the result of comparing the processing of coherent phrases ("29" …

"inches") with the processing of items that will not be conveying any kind of message-level, unified information (as the probe items in the other tasks are simply asking a question about the prior number). More importantly, although the exact source of this effect is unclear without further empirical work, we take this pattern as a clear indication that different sorts of cognitive mechanisms were engaged by our tasks (which otherwise patterned similarly).

In sum, the very small, but reliable, effect of repetition and its unusual scalp distribution both are consistent with the idea that there might be a precisely tuned and specialized conceptual representation of semantics for Arabic numerals. Whereas the semantics accessed during word reading and picture viewing can be relatively more broad in scope, perhaps numerals elicit a more focal and specific kind of meaning, resulting in both a novel scalp distribution and a smaller effect of repetition priming. Arabic numerals operate in a unique space in our long term memories: we view them often enough to be familiar with them, but their meaning tends to be more specific than that elicited both by pictures (whose semantics can be ambiguous) and by words (e.g., reading "mop" engages the semantics not just of mops but also of similar-looking words "pop" and "top" and the semantics associated with them as well; Laszlo & Federmeier, 2009). In contrast, it may be to our benefit to both associate Arabic numerals with very specific semantics (i.e., number words and their values) and to restrict the automatic spreading of semantic activations across numerals more than is done for written words.

Figures

Figure 4.1

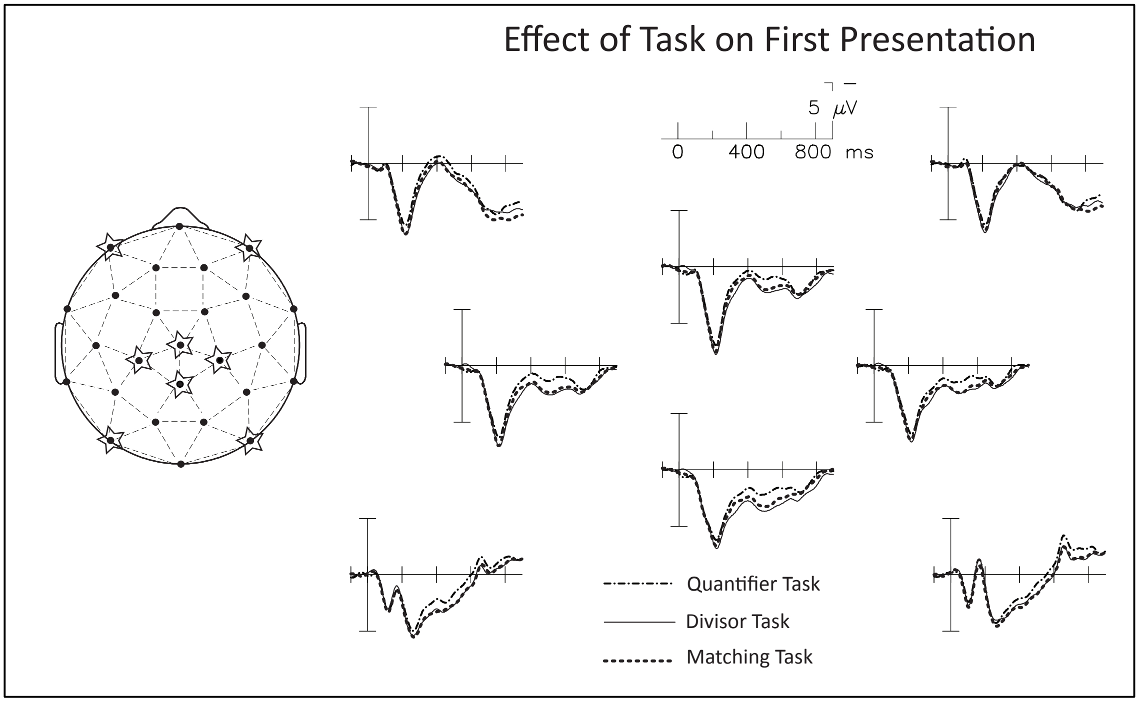

Figure 4.1 *Left:* Schematic of electrode channel configuration, with stars indicating the plotted channels. *Right:* Response to the first presentation of Arabic numerals in each task (quantifier, divisor, and matching). There is a more negative response to these numerals in the quantifier task than in either the divisor or matching tasks.

Figure 4.2

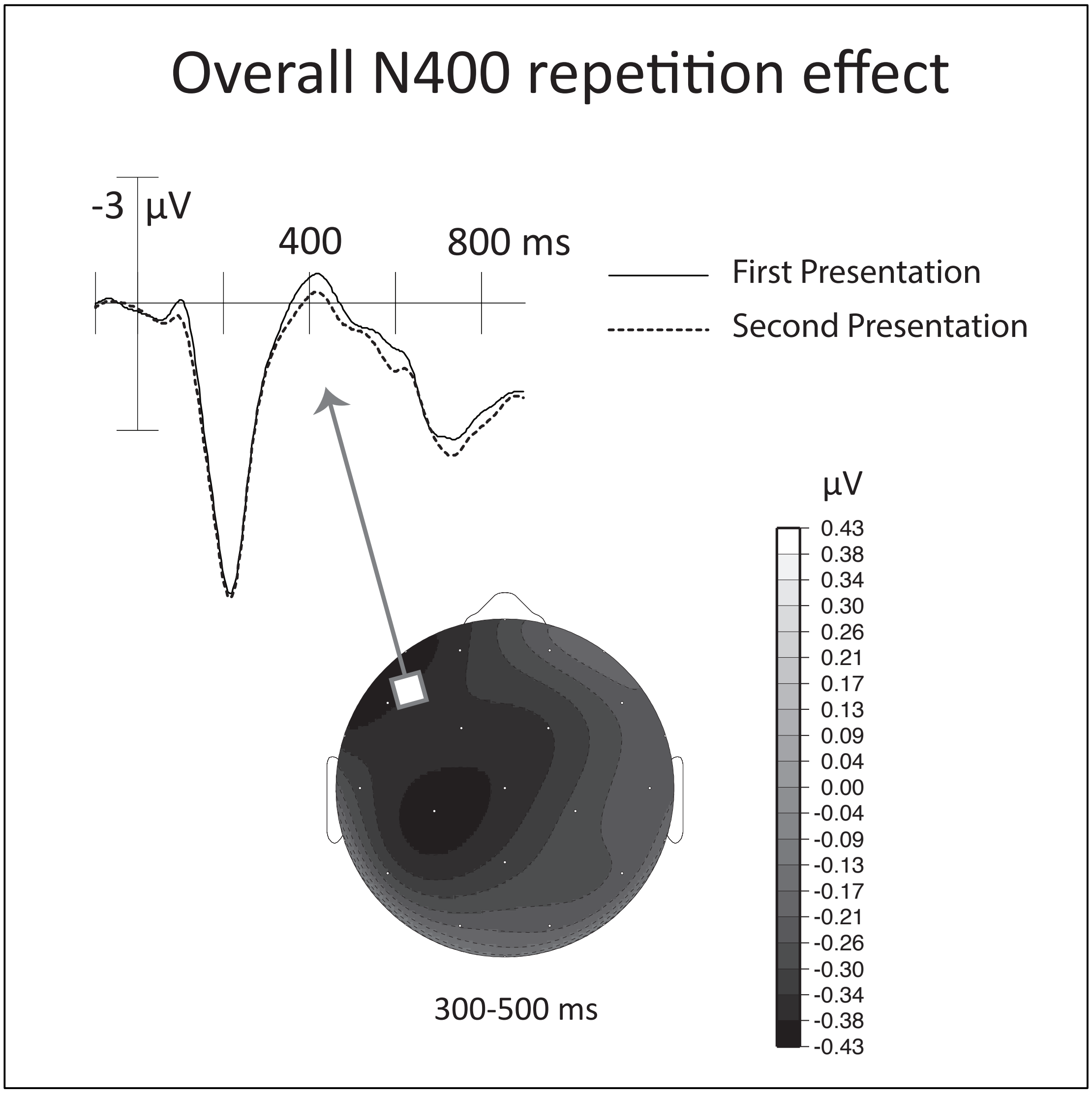

Figure 4.2

Top: The effect of repetition of numerals across all tasks is plotted, where it can be seen that numerals are more positive on their second presentation compared to their first. *Bottom:* Graphical representation of the topographical distribution of the effect of repetition on the N400 (measured 300 – 500 ms post-stimulus as the numerical subtraction of first presentation from second presentation). The darker areas correspond to a larger repetition effect.

Figure 4.3

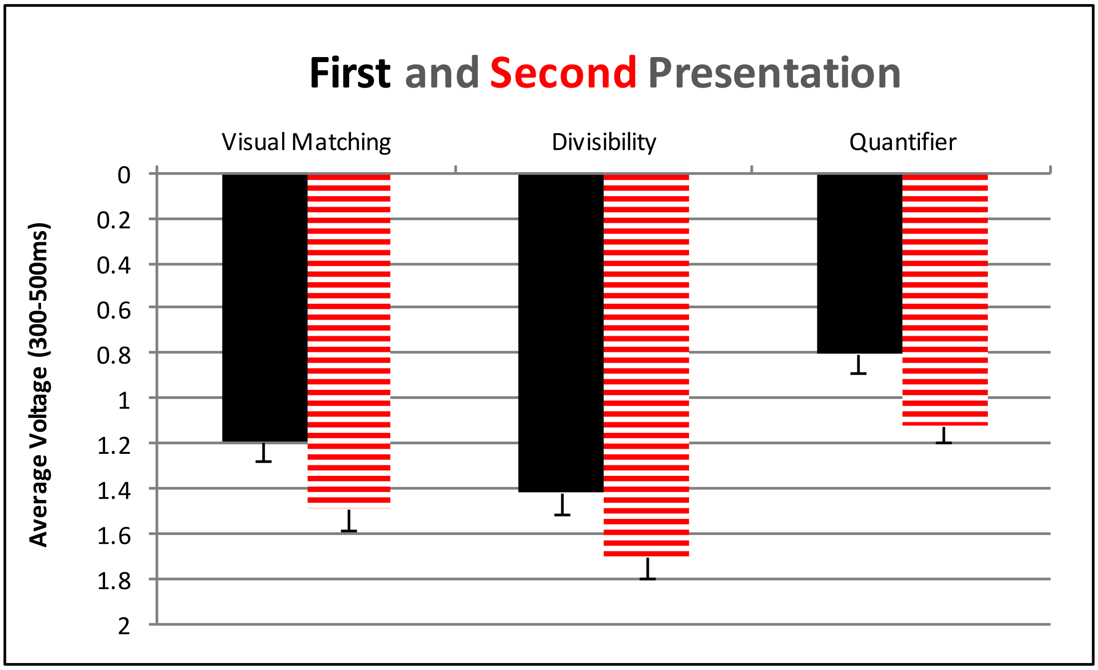

Figure 4.3 The N400 repetition effect measured in each task is presented, showing more positive responses for second presentation (striped red) compared with first (black). The average voltage from 21 electrode channels measured over 300 - 500 milliseconds post-stimulus onset in each task is reported. Positive is plotted down for consistency with ERP plotting customs.

Figure 4.4

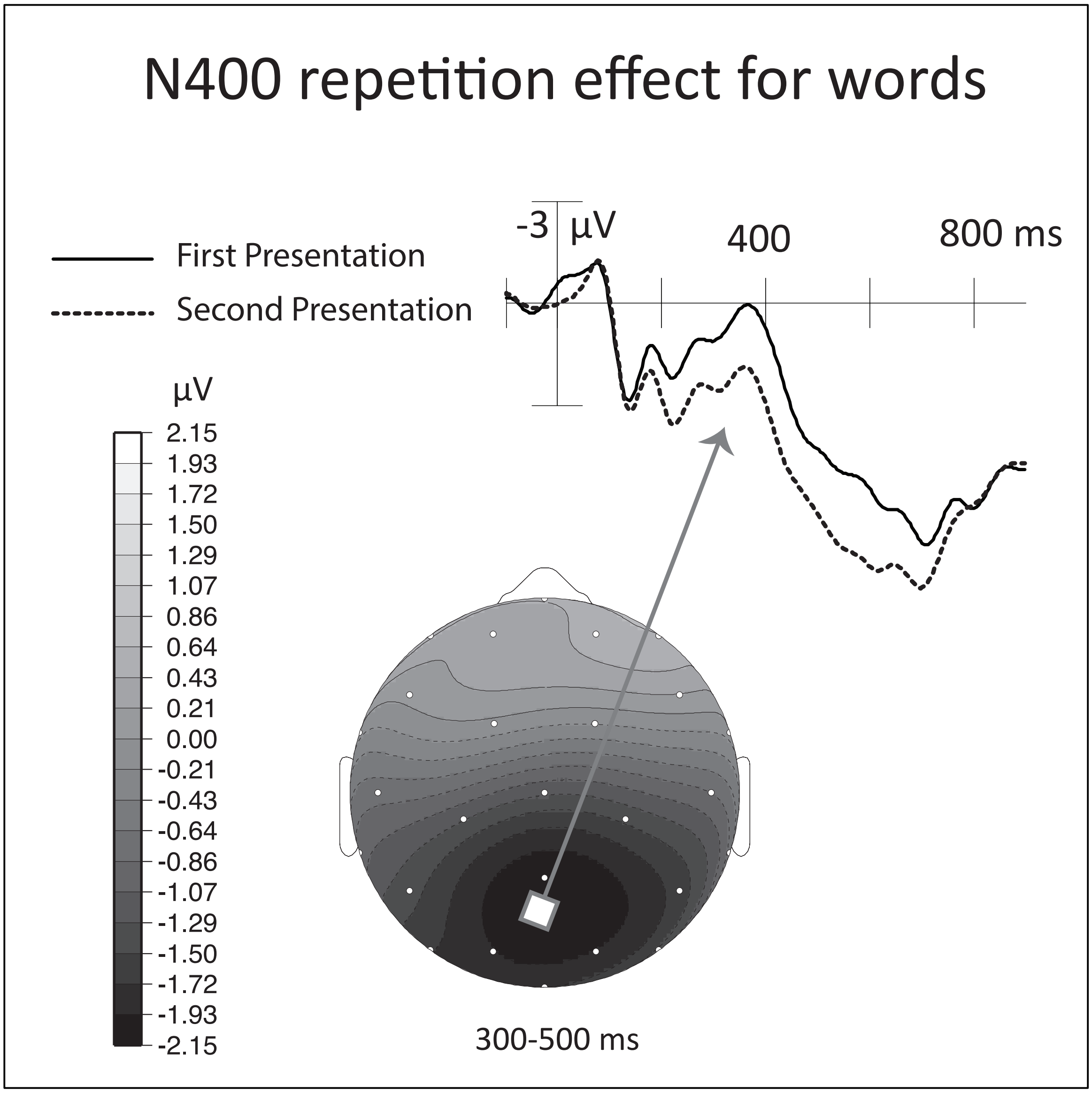

Figure 4.4

Top: Response to the repetition of probe quantifier words. *Bottom:* Graphical representation of the topographical distribution of the effect of repetition on the N400 (measured 300 – 500 ms post-stimulus as the numerical subtraction of first presentation from second presentation). The darker areas correspond to a larger repetition effect.

Figure 4.5

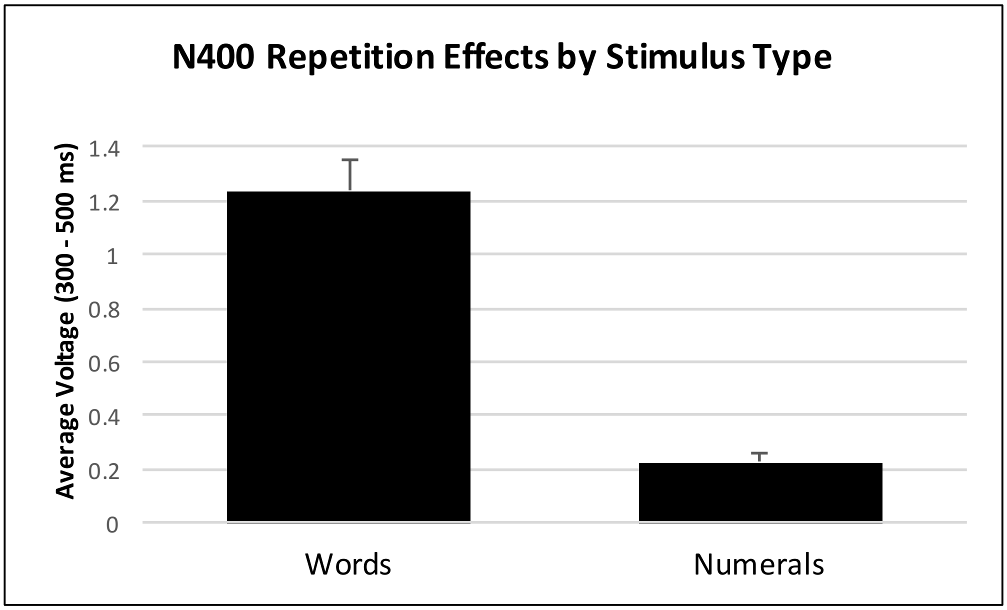

Figure 4.5 The N400 repetition effect measured for words and double-digit Arabic numerals. Plotted is the difference between first and second presentation of items, as derived from the average voltage of 15 centroposterior electrode channels measured over 300-500 milliseconds post-stimulus onset.

CHAPTER 5

Experiment 4: Is your favorite number special? Evidence for item-level differences in retrieval of information from numerals

Introduction

Experiment 3 showed that, although N400 repetition effects can be obtained for Arabic numerals, such effects are smaller than those obtained for other types of meaningful stimuli, and their size seems largely unaffected by the amount of conceptual processing required by the task. The N400 repetition effects to Arabic numerals also differed from those seen to words in their scalp distribution. These differences raise important questions about how these numerals, as compared with other stimulus types, access meaningful representations in long term memory. Here, therefore, we ask more directly what happens when people are encouraged to process numerals for meaning in a manner more similar to words, and, in that context, whether people's experience with (individual) numbers matters. First, it is useful to examine the range of repetition effects that have been obtained in the literature thus far and what variables influence those effects.

One of the most salient features of the repetition effect obtained to numbers in Experiment 3 was its small size. Generally, N400 repetition effect sizes for a given stimulus type are dependent on a number of factors. Two relatively straightforward modulators of repetition effect size are retention duration (fewer intervening items or less delay between repetitions, larger effects of repetition) and number of repetitions (larger effects for the first repetition than subsequent ones). Other modulators of repetition effect size are relatively more subjective to manipulate, and they include task, with larger effects for tasks involving deeper engagement with the critical items, and type of stimulus, with more conceptually rich stimuli (e.g., words) eliciting larger repetition effects than conceptually impoverished stimuli (e.g., randomly connected lines) (Nagy & Rugg, 1989; Rugg, Furda, & Lorist, 1988; Rugg & Nagy, 1989; Rugg, 1990; Van Petten & Senkfor, 1996).

The most straightforward account of N400 repetition effects is that they are due to a reduction in the amount of *new* semantic activity that occurs as a result of prior semantic long-term memory access. That is, the semantic state established with the first presentation of an item persists to some degree over time (e.g., because semantic features remain active or, on a recent computational model of N400 repetition effects, because the influence of inhibitory connectivity persists, making it harder to reactivate the same semantics; Laszlo & Armstrong, 2014). With time, these influences from the prior presentation decay, making the processing of repetitions at increasingly long intervals more and more like that of first presentation. This also explains why additional repetitions are less effective at inducing additional facilitation: it might only take one exposure to a given stimulus to make strong contact with long-term memory representations, such that there is little change of state with additional exposures.

Under this view, the reason that conceptually driven tasks are more effective at eliciting repetition effects is because they encourage readers to fully activate and refine the semantic features of the incoming stimulus (including processing that occurs *after* the N400 at first presentation). In the case of words, semantic access seems to be relatively obligatory, such that repetition effects are obtained even when words are presented in the context of detection tasks targeting physical properties of the word; however, these effects are more substantial in detection tasks targeting semantic category membership (e.g., Rugg, Furda, & Lorist, 1988). For other classes of stimuli, such as unpronounceable strings of letters ("illegal strings") like "TXM" and unfamiliar geometric shapes, effects of task on repetition effects are even more striking. In many task contexts, repetition effects for such stimulus types are small or not detectable (Rugg & Nagy, 1987; Voss & Paller, 2009b; Voss & Paller, 2010). For example, when participants are asked to monitor an input stream for a category of items (such as proper names), illegal strings fail to elicit N400 repetition effects (Laszlo & Federmeier, 2007), presumably because these items are easily dismissed as task irrelevant and thus not processed deeply. However, when task demands require that the meaningfulness of these strings at least be considered – i.e., in a modified lexical decision task with "yes" responses to meaningful acronyms like "CBS" and "DVD" as well as to words -- then such strings also elicit N400 repetition effects (Laszlo, Stites, & Federmeier, 2012). In sum, when stimuli can be immediately dismissed as task-irrelevant, a decision can be made without full access to associated long-term memories, and, as a consequence, smaller (or nonsignificant) N400 repetition effects are obtained.

Experiment 3 provided evidence that Arabic numerals are an interesting class of stimuli that elicit some long-term semantic memory processing under a range of task conditions, but do so in a manner that results in a notably smaller amount of repetition-based N400 facilitation than for more broadly meaningful stimuli like words. Importantly, although manipulations of task failed to modulate repetition effects for numerals as a collective class of stimuli, there is also evidence that repetition effects can vary on the level of *items* within a stimulus class. Specifically, one way that researchers have succeeded in finding evidence that otherwise low-meaningfulness stimuli can elicit stable conceptual representations (and thus elicit repetition effects) is through sorting the items by their subjective meaningfulness to a given participant. This is possible because even among classes of stimuli that have generally been associated with lower levels of rich conceptual information, there can still be items that a given individual finds to be more or less meaningful to them. For example, among illegal strings of letters, some are meaningful – e.g., are familiar acronyms – and others are not, and, indeed, the effects obtained by Laszlo & Federmeier (2007) were based on person-specific item-level meaningfulness, as some acronyms were not familiar to individual people (and thus elicited responses like those to non-acronym illegal strings) and some illegal strings were idiosyncratically familiar (for example, BWHS might be the acronym for a particular person's high school), and consequently elicited acronym-like repetition effects. Thus, an individual participant's experience with an item can influence the magnitude of his/her N400 repetition effects to that item.

Similar studies of the flexibility of our conceptual engagement with relatively unfamiliar-looking items have been performed with item classes outside of letters. For example, novel geometric shapes can sometimes appear to be familiar (e.g., what appears to be a random "squiggle" to one person might appear to be a sketch of a jumping deer to another person). If individuals are not told about this possibility, and are simply asked to memorize these items or to judge them for a low-level visual feature, there is no evidence of semantic processing (i.e., no N400 repetition effects; Van Petten & Senkfor, 1996; Voss, Schendan, & Paller, 2010; Voss & Paller, 2009b). This is similar to how meaningless illegal strings do not elicit repetition effects when surface-level assessments of them ("not a possible word") render their meaningfulness task-irrelevant. However, when individuals are encouraged to think about the possible meaningfulness of novel shapes because the task is to rate them as personally familiar or unfamiliar, then evidence of conceptual priming emerges, selectively for the shapes that individuals rated as familiar (Voss et al., 2010).

It is important to note that, even in the same task ("rate for meaningfulness"), strings of text and novel shapes differ in their pattern of N400 repetition effects. Under these processing conditions, letter strings elicit N400 repetition effects regardless of their meaningfulness ratings, whereas shapes only elicit N400 repetition effects if the participant had identified them as familiar/meaningful (Laszlo et al., 2012; Voss et al., 2010; c.f., Voss, Lucas, & Paller, 2010, with a longer retention interval and more intervening items, which led to only wordforms rated as familiar eliciting facilitation on the N400). In general, this work with semantic processing of letter strings and geometric shapes suggests a natural readiness for conceptual engagement with *any* class of stimuli – and the strength of this engagement depends on readers' subjective judgment of their meaningfulness.

This literature raises the interesting possibility that, although manipulations of conceptual engagement with numerals as a *class* was ineffective at modulating the size of N400 repetition effects (in Experiment 3), there might be differences in meaningfulness at the level of *item* – that is, some Arabic numerals might themselves carry broader meaningfulness than others. Such a possibility could then be uncovered by participants' subjective judgments. In order to address this question, in the present study we had participants read the same double-digit Arabic numbers as in Experiment 3, while rating them as either familiar (i.e., meaningful) or not. We then examined N400 repetition effects for these items as a function of participants' own perception of them as familiar or unfamiliar.

In general, this approach of self-rated meaningfulness rests on a relatively new way to think about how semantic access for numerals unfolds. The majority of the literature treats the learned association between numerals and specific values and/or numerosities as their most important feature (i.e., "16" is a symbol that is highly associated with the number word "sixteen", and the primary meaning of "16" is a *magnitude* of 16 units), given their critical role in arithmetic and mathematical processing. Debates over the meaning of numerals have tended to circle whether or not quantity information is automatically activated from numerals (i.e., digits) (Henik & Tzelgov, 1982; Pavese & Umiltà, 1998; Ganor-Stern, Tzelgov, & Ellenbogen, 2007; Cohen, 2009; Lyons, Ansari, & Beilock, 2012; Goldfarb, Henik, Rubinsten, Bloch-David, & Gertner, 2011), whether other aspects of digits influence how they are accessed (e.g., physical similarity to each other) (Garcea-Orza, Perea, Mallouh, & Carreiras, 2012), how much of their meaning is structured through a representation resembling a number line (Umiltà, Priftis, & Zorzi, 2009; Santens & Gevers, 2008; Fias, van Dijck, & Gevers, 2011), and whether or not the representation of numerosity engaged by Arabic numerals is shared in an abstract form with the representation of numerosity activated by number words and by non-symbolic representations of number (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Libertus, Woldorff, & Brannon, 2007; Cohen Kadosh & Walsh, 2009).

With respect to these debates, the results from Experiment 3 were broadly consistent with the notion that *some* kind of meaning is being accessed relatively automatically when numerals are read, in the sense that effects emerged on what would seem to be the N400, which is a relatively automatic index of meaning access (i.e., N400 responses are elicited to potentially meaningful stimuli essentially whenever they are encountered, even when participants are unable to report having perceived the stimuli; Vogel, Luck, & Shapiro, 1998). Additionally, the results were consistent with the possibility that the specific meaning could be numerical in nature, since the distribution of the effect is different from words – and, thus, the nature/source of the semantics could be different. Finally, the consistently small magnitude of the effect could be consistent with the idea that the set of information being accessed from memory is fairly *restricted* (e.g., limited to size or number value), at least relative to the concepts associated with other stimulus types like words. What the present experiment proposes to uncover, then, is a relatively independent possibility that, under circumstances wherein participants are encouraged to think about the meaning of numbers more broadly, the semantic network that is activated by numerals could additionally include other types of (more word-like) meaningfulness, beyond the more specific concepts that they otherwise seem to bring to mind.

The most direct form of evidence for this would be if items subjectively rated as familiar elicited substantially larger N400 repetition effects than items rated as unfamiliar. If, in contrast, items rated as familiar and items rated as unfamiliar elicit similarly small N400 repetition effects, then that would suggest that our relatively automatic semantic access for numerals is strikingly consistent across items (as well as across task, as seen in Experiment 3). In sum, whereas in Experiment 3 we attempted find evidence that different tasks could modulate the amount of semantic access a class of items received, in the present experiment we attempt to find evidence for differences in semantic access across items within the same (familiarity judgment) task.

Another feature of the N400 repetition effect that could reveal a change in what people are accessing when they view Arabic numerals (rather than just how much semantic content they are accessing) is the scalp topography of the effect itself. Notably, N400 repetition effects for more broadly meaningful pictures and letter strings tend to have centroparietal distributions. In Experiment 3, we observed a different distribution – more frontal and left-lateralized. It is possible that if we ask participants to think about the broader potential meaningfulness of these numerals, then we might find a more canonical (word-like) distribution of the N400 repetition effect. This would suggest that numerals can elicit access the same kind of broad meaningfulness as other conceptually rich items like words and pictures.

Although N400 repetition effect patterns are a primary focus of this experiment, there are other ERP responses that may also be affected by our manipulation. These additional effects offer the opportunity to further assess how numbers may be treated similarly or differently from other potentially meaningful stimuli. Specifically, in repetition paradigms, beyond effects on the N400 (which are thought to reflect relatively automatic and effortless semantic access), there are also frequent reports of differences on late positivities (conventionally referred to as a "late positive complex" or LPC), which are instead thought to reflect more explicit and conscious recollective processing of stimuli (e.g., Paller & Kutas, 1992; Paller, Kutas, & McIsaac, 1995; Voss & Paller, 2009a). The strongest evidence for a distinction between N400 and LPC repetition effects comes from patients with hippocampal damage, who have severe impairments in the ability to form and explicitly recall new memories of items and events, although they often retain the ability to learn subconsciously (e.g., there is evidence for implicit priming in amnesic patients) (e.g., Warrington & Weiskrantz, 1970; for review, see Schacter, 1987). In these hippocampal patients, repetition effects can be obtained on the N400, but LPC effects are eliminated (Olichney et al., 2000; Duzel, Vargha-Khadem, Heinze, & Mishkin, 2001). Also consistent with the interpretation of LPC effects as being dependent on intact hippocampal function and associated with declarative memory, larger LPC effects are typically associated with better behavioral performance (e.g., more successful recollection of the source of a given memory) in recognition tasks (Curran & Cleary, 2003; Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997; Woodruff, Hayama, & Rugg, 2006; Woroch & Gonsalves, 2010).

These LPC effects have been studied both in incidental repetition paradigms – wherein explicit old/new categorization is not required – as well as in paradigms in which the second presentation of the item takes place in the context of a memory test (i.e., distinguishing between old and new items). In explicit testing paradigms, larger LPCs are typically obtained for correctly recognized old items compared to new or missed (incorrectly rejected old) items (e.g., Van

Petten & Senkfor, 1996; Voss & Paller, 2009a,b). Interestingly, the same geometric shapes that failed to elicit N400 repetition effects when they were judged to be unfamiliar also failed to elicit these LPC recollection effects, suggesting that the presence of the LPC is not just due to remembering that the item was previously presented, but reflects explicit retrieval of details about that item (Voss, Schendan, & Paller, 2010).

In the present experiment, participants are rating their own subjective familiarity with each numeral, and it seems likely that, in order to make these judgments about the personal relevance of a numeral due to prior experiences, recollective processes necessarily occur. In Experiment 3, which did not obtain any LPC effects, processes of recollection (and decision-making in general) were purposefully delayed until the test probe occurred. Thus, even though previously these stimuli did not elicit late positivities, it is possible that they might in the current paradigm, particularly since participants are being asked to engage in explicit retrieval.

On the basis of the prior literature, then, if LPC effects for numerals are obtained, we expect to see enhanced LPC responses (more positivity, beginning around 500 ms and continuing for a few hundred milliseconds) for repeated items (which are similar to "old" items in explicit memory paradigms). A key novel question, then, is whether this potential LPC enhancement will also be selective (or at least enhanced) to numerals that were rated as subjectively familiar. Such a finding would be consistent with an interpretation that our long term memories for the subset of Arabic numerals that are uniquely meaningful to us can be deliberately retrieved and impact our subsequent reading of them. An additional possibility is that, on the measure of the N400, no difference between numerals rated familiar and unfamiliar might emerge, but instead an interaction might occur selectively on the LPC. This was the case in a study involving complex visual shapes (Voss & Paller, 2009b), in which stimuli were too conceptually impoverished to elicit facilitated N400 responses but still elicited enhanced LPCs during recognition. Therefore, failures to find evidence that Arabic numerals are processed for their distinctive amount of meaningfulness on the N400 would not immediately indicate that there is *no* additional information about aspects of their meaning (e.g., to learned associations beyond numeracy) in long term memory, only that accessing this extra meaning is not fast and obligatory.

In general, the pattern of N400 and LPC repetition effects for numerals rated as familiar and unfamiliar will help us better understand how people engage with the meaning and long-term memory representation of these numerals outside of their utilization as a representation of specific numerical value.

Methods

Participants

Thirty-two participants were included for analysis in this study. All were right-handed native monolingual English speakers (half female) with normal or corrected-to-normal vision, no early exposure to a second language, history of brain trauma, or current use of psychoactive medications. Participants' average age was 21 (range: 18 – 29), their mean laterality quotient on the Edinburgh handedness inventory (Oldfield, 1971) was 0.82 (range: 0.43 - 1.00, with 1.0 indicating strongly right-handed and -1.0 strongly left-handed), and 9 reported having a left-handed biological family member. Informed consent was received from all participants in accordance with the Institutional Review Board at the University of Illinois.

Stimuli

The critical stimuli were the same as from Experiment 3 (double-digit numbers, repeating after a 2-3 intervening trials), and each participant saw one of three list orders.

Procedure

The basic procedure was the same as in Experiment 3. The critical numbers were presented for 500ms, followed by a blank data collection period of 800ms, after which subjects were prompted by a question mark to enter their response categorizing the item as Familiar or Unfamiliar to them. Once their response to a numeral was registered, they proceeded to the next trial. For a numeral to be categorized as Familiar, participants were instructed to use a liberal and subjective definition of either having specific life experiences with the numeral and/or more generally having a high degree of personal familiarity with it (that did not necessarily need to be linked to a concrete reason). This relatively open description was necessary to generate enough endorsements of meaningfulness to have sufficient trial counts for each subject: under more restrictive criteria, most subjects would not endorse many numerals as meaningful.

All subjects were provided the following specific examples of how to think about concrete meaningfulness: apartment numbers, people's ages, dates/years, and athlete jersey numbers. Responses were monitored during the experiment to ensure compliance with instructions, and some participants had to be encouraged to be more liberal with their verifications (i.e., they started out rating every numeral as Unfamiliar). Response hand was counterbalanced with participant sex and stimulus list (with half responding Familiar with their right hand and Unfamiliar with their left hand).

EEG Recording

Brainwaves were recorded using 26 silver/silver-chloride electrodes mounted in an elastic cap. All electrode impedances were kept below 5kΩ, and electrodes were evenly distributed over the scalp (see Figure 2.1 for the configuration). The data were referenced on-line to the left mastoid and re-referenced offline to an average of the right and left mastoids. A separate frontal electrode acted as ground. The vertical electrooculogram (EOG) signal was monitored with an electrode just below the eye (on the infraorbital ridge) and horizontal EOG was monitored with electrodes placed on the outer canthus of each eye. The EEG signal was sampled at a 250 Hz and was subjected to an analog bandpass of 0.02 to 100 Hz during online amplification by Sensorium amplifiers.

Raw waveforms were assessed for inclusion on a trial-by-trial basis with artifact thresholds separately calibrated by visual inspection for each subject. Trials were excluded from averaging if they included blinks, movement artifacts, signal drift, blocking, or a horizontal eye gaze movement. After artifact rejection, for items rated familiar consistently (endorsed on both first and second instance), there were 26 trials on average (range: 8 – 41; median: 27), and for items rated unfamiliar consistently, there were 42 trials on average (range: 14 – 70; median: 39).

Epochs of EEG data for each trial were taken from 100ms prior to item onset until 920ms after item onset, and the baseline acquired over the 100ms preceding the onset of each trial was subtracted prior to averaging. ERP mean amplitudes were measured after application of a digital bandpass filter of 0.1 to 20 Hz. For all tests with more than 1 degree of freedom in the numerator, the Huynh-Feldt corrected P-value is reported. In all of the results below, interactions and effects of electrode-related factors are only reported for those of theoretical relevance.

Results

Behavior

Familiarity ratings

Although most numbers had a bias towards overall familiarity/unfamiliarity, there was across-subject disagreement: the top 30% of numbers consistently rated as Familiar were still rated as Unfamiliar more than 25% of the time. All numbers received mixed ratings across participants. There was a tendency for numbers that were more consistently rated as Familiar to be lower in value. [4]

Event-related potentials

N400 timewindow

Analysis of the latency of the N400, as measured by identifying the peak of the negativity in the timewindow from 250 to 500ms, found that it was maximum at 372ms (SE = 1.3ms) post-stimulus onset. There were no effects of Familiarity or Repetition on its latency. In the analyses that follow, N400 amplitude was measured from 300 – 500ms (as in Expt. 3).

To assess effects of Repetition and Familiarity on the N400 and to characterize the scalp distribution of any such effects, we performed an omnibus ANOVA with 2 levels of Familiarity (Familiar, Unfamiliar), 2 levels of Repetition (First Presentation, Second Presentation), 2 levels of Hemisphere (Left Hemisphere, Right Hemisphere), 2 levels of Laterality (Lateral, Midline), and 4 levels of Anteriority (from front to back) on the 16 electrode sites that could be classified along all three dimensions. This analysis revealed a significant main effect of Familiarity

---

[4] This is to be expected given that numerals were not an experimentally generated random set of stimuli, but instead were stimuli that participants had a lifetime of experience using, and lower numbers are more frequent in the world. However, the average value of the top-third most consistently rated Familiar numbers (38) was still higher than would be expected if the numbers had simply been ranked by value (24.5). Similarly, the average value of numbers most consistently rated Unfamiliar (64) was lower than what would be expected by a simple ranking of the 30 highest-value numbers in the sample (75.5).

($F(1,31) = 14.92$, $p < 0.001$), which interacted with the distributional variables of Laterality ($F(1,31) = 12.32$, $p < 0.01$) and Anteriority ($F(3,93) = 10.93$, $p < 0.01$), as well as with Laterality and Anteriority together ($F(3,93) = 4.0$, $p < 0.05$). There was no main effect of Repetition ($F(1,31) = 1.67$, $p > 0.05$), but there were interactions between Repetition and the distributional variables of Laterality ($F(1,31) = 5.57$, $p < 0.05$) and Anteriority ($F(3,93) = 4.38$, $p < 0.05$). Finally, there was an interaction between Familiarity, Repetition, and the three distributional factors ($F(3,93) = 3.06$, $p < 0.05$).

The reported interactions between critical factors and the distributional factors reflected effects were generally larger at medial sites (generating effects of Laterality) and that there were differences in how these effects patterned across frontal and posterior channels (generating effects of Anteriority). Such effects of Laterality are not surprising, given the use of a mastoid references which results in smaller effects at temporal (lateral) sites. However, effects of Anteriority are likely more substantive. Therefore, to better characterize the effects of Familiarity and Repetition across levels of Anteriority, separate analyses of these factors were performed with the 11 frontal channels and with 15 posterior channels.

In frontal channels, there was only a main effect of Familiarity ($F(1,31) = 18.23$, $p < 0.001$), and no effect of Repetition ($F(1,31) = 0.01$, $p > 0.50$) or interaction between Repetition and Familiarity ($F(1,31) = 0.46$, $p > 0.50$). The main effect of Familiarity was due a more positive response to items rated as Familiar (3.64μV, SE = 0.15) compared to items rated as Unfamiliar (2.30μV, SE = 0.12).

In posterior channels, however, there were both main effects of Familiarity ($F(1,31) = 9.33$, $p < 0.01$) and Repetition ($F(1,31) = 7.67$, $p < 0.01$). There was no interaction between Familiarity and Repetition ($F(1,31) = 0.82$, $p > 0.50$). The main effect of Familiarity was due to a more positive response to items rated as Familiar (5.22µV, SE = 0.15) compared to items rated as Unfamiliar (4.50µV, SE = 0.10), and the main effect of Repetition was due to Second Presentation items (5.28, SE = 0.11) being more positive than First Presentation items (4.53, SE = 0.10).

In sum, both frontal and posterior channels had an effect of Familiarity that was unrelated to Repetition, with more positivity (or less negativity) for items rated as Familiar than items rated as Unfamiliar, whereas posterior channels revealed an additional effect of Repetition that was unrelated to Familiarity (with facilitation for repeated items). The leftmost plots in Figure 5.3 display topographical maps of the distribution of the N400 repetition effect, separately for items rated Familiar and Unfamiliar. Raw waveforms of the N400 repetition effect from two representative channels are plotted in Figure 5.2.

LPC timewindow

Analysis of the latency of the late positivity, as measured by identifying the maximum of the positivity in the timewindow from 500 to 900ms, found that it was maximum at 611ms (SE = 2.1ms) post-stimulus onset. There were no effects of Familiarity or Repetition on its latency. In the analyses that follow, LPC amplitude was measured from 500 – 800ms.

To assess effects of Repetition and Familiarity on the LPC and to characterize its scalp distribution, we performed an omnibus ANOVA with 2 levels of Familiarity (Familiar, Unfamiliar), 2 levels of Repetition (First Presentation, Second Presentation), 2 levels of Hemisphere (Left Hemisphere, Right Hemisphere), 2 levels of Laterality (Lateral, Midline), and 4 levels of Anteriority (from front to back). This analysis revealed a main effect of Familiarity ($F(1,31) = 29.73$, $p < 0.001$), which interacted with the distributional variable of Laterality

($F(1,31) = 16.56$, $p < 0.001$), and with both Laterality and Anteriority ($F(3,93) = 3.81$, $p < 0.05$). There was no main effect of Repetition ($F(1,31) = 0.29$, $p > 0.50$), but this factor interacted with the distributional variable of Anteriority ($F(3,93) = 16.27$, $p < 0.001$), as well as with Laterality and Anteriority ($F(3,93) = 3.23$, $p < 0.05$) and Hemisphere and Laterality ($F(1,31) = 2.07$, $p < 0.05$). Additionally, Repetition and Familiarity interacted with the distributional variable of Anteriority ($F(3,93) = 6.60$, $p < 0.05$), as well as all three distributional factors ($F(3,93) = 2.83$, $p < 0.05$).

To examine the effects of Familiarity and Repetition and their interactions with Anteriority, separate analyses were performed over the frontal 11 channels and the posterior 15 channels. Over posterior channels, there were main effects of Familiarity ($F(1,31) = 21.32$, $p < 0.001$) and Repetition ($F(1,31) = 5.53$, $p < 0.05$), and an interaction between Familiarity and Repetition ($F(1,31) = 5.53$, $p < 0.05$). Over frontal channels, there was only a main effect of Familiarity ($F(1,31) = 31.87$, $p < 0.001$). Thus, in the omnibus ANOVA, the interactions between Familiarity and Repetition with Anteriority largely reflected that, in posterior channels, there was an interaction between Familiarity and Repetition, whereas in frontal channels, there was a main effect of Familiarity that was not impacted by Repetition.

Planned comparisons of the effect of Repetition on Familiar and Unfamiliar items were performed over the posterior channels, where LPC effects typically emerge. For items that were rated Unfamiliar, there was no effect of Repetition ($F(1,31) = 0.53$, $p > 0.10$). However, for items that were rated as Familiar, there was an effect of Repetition ($F(1,31) = 7.90$, $p < 0.01$), with Familiar Second Presentation (repeated) items being more positive (6.07μV, SE = 0.17) than Familiar First Presentation (unrepeated) items (4.82μV, SE = 0.18). The rightmost plots in Figure 5.3 display topographical maps of the distribution of the LPC effect, separately for items rated Familiar and Unfamiliar. Raw waveforms of the LPC effect from two representative channels are plotted in Figure 5.2.

The same analysis performed over the frontal 11 channels found no effect of Repetition either for items rated Familiar ($F(1,31) = 1.95$, $p > 0.10$) or items rated Unfamiliar ($F(1,31) = 1.02$, $p > 0.10$). Over these channels, there was (as in the N400 timewindow), a main effect ($F(1,31) = 22.38$, $p < 0.001$) wherein items rated as Familiar were more positive / less negative (5.34µV, SE = 0.21) than were items rated as Unfamiliar (3.81µV, SE = 0.17). Figure 5.1 displays this frontal effect of Familiarity.

Discussion

Our primary goals were to assess how subjective ratings of familiarity of double-digit Arabic numerals affected participants' (1) conceptual processing of them (indexed by N400 repetition effects), and (2) explicit retrieval of detailed information about them (indexed by LPC repetition effects). Importantly, these effects were assessed within-subject (based on their subjective ratings of personal experience with Arabic numerals), so that, across participants, any particular numeral varied in whether it was categorized as familiar or unfamiliar.

Our findings for the effect of repetition in the N400 timewindow are as follows. First, we found a significant N400 repetition effect. Strikingly, this effect had the canonical (centroparietal) distribution, suggesting that we succeeded in shifting the kind of semantics people engage with when reading numerals to a more word and picture-like representation. Interestingly, this effect was also numerically larger than that seen in Experiment 3 (around 0.75µV versus the prior 0.30µV). In general, the features of the N400 effect we observed (i.e., its canonical distribution and at least numerically more substantial size) seem to collectively indicate that the demands of the familiarity judgment task shifted the nature and/or amount of semantic information obtained from numerals. However, critically, there was not a difference in this effect across familiarity.

Numerals rated familiar and numerals rated unfamiliar elicited indistinguishable (small) N400 repetition effects (see Figure 5.2 for these waveforms). This replicates findings from a similar task with words and letter strings (Laszlo et al., 2012). There, even unpronounceable strings of letters rated as unfamiliar ("TXQ") elicited N400 repetition effects when participants were judging whether or not they were meaningful (i.e., whether or not they were an acronym). Further, this is not what has been observed when participants rated novel shapes (squiggles) – there, only shapes rated as familiar elicited N400 repetition effects (Voss et al., 2010). That is, unlike truly pre-experimentally novel geometric shapes, double-digit numbers were similarly processed for their meaningfulness regardless of how participants eventually subjectively rated them. This might indicate that having long-term, pre-experimental experience extracting meaning from a stimulus class renders even those items rated as unfamiliar more meaningful than items similarly that are similarly rated as unfamiliar but for which participants have no pre-experimental experience. The mechanisms resulting in this outcome might differ across stimulus class, however: whereas strings of letters might be more inherently meaningful (than novel shapes) due to their confusability with real words and acronyms, a different process is likely unfolding when encountering Arabic numerals and accessing their semantics.

Importantly, under many assumptions of Arabic numeral reading, some kind of numerical magnitude or value information is obligatorily accessed every time a numeral is encountered, in a mostly task-independent way (and whether these representations are for the exact number value or for a more abstract conceptual numerosity is contentious). Under this view, it would be expected that even numbers being rated as unfamiliar should convey some kind of stable conceptual magnitude/value representation to the viewer, which would suggest there should not be a huge difference in N400 repetition effects for items rated familiar and unfamiliar. Indeed, this is what we observed. But does this necessarily mean that the N400 repetition effect we found for Arabic numerals subjectively rated as unfamiliar reflects facilitation of the processing of numeracy-related concepts?

We suspect that the shift of the N400 effect from the unusual frontal-left scalp distribution in Experiment 3 (where the semantic representations being engaged might have been number value or numeracy alone) to the more canonical centroposterior scalp distribution in the present experiment might argue against such a consistent uniformity in the precise type of semantics being accessed. At the very least, something about the neural source of the N400 repetition effects across these experiments differed, and the sources responsible for repetition effects in the present experiment seem more similar to those that typically give rise to N400 repetition effects in more richly meaningful contexts than did the sources of the effects observed in Experiment 3.

One possibility is that concepts related to numeracy were indeed uniformly being activated, but *additional* semantic activations were encouraged by the current experimental paradigm that both caused the effect to be somewhat larger and to shift in distribution into posterior regions. The size of the effect over left frontocentral sites that were the focus of the prior experiment remains similar here (first presentation: 2.79 μV, SE = 0.19; second presentation: 3.15 μV, SE =0.20), making it especially tempting to interpret the present effect as at least somewhat of an augmentation of the prior effect. While this is possible, having different participants across studies precludes any strong interpretation of this kind. Further, any apparent numerical facilitation over these sites in the present study could simply reflect residual spreading of the N400 effect away from its posterior maximum (see Figure 5.3). A within-subject design with more a targeted approach to addressing this issue might be able to better disentangle these possibilities.

Regardless of what the precise source of the semantic activity across these studies was, it is clear that the N400 repetition effect itself was not affected by ratings of subjective familiarity. However, there was an effect of familiarity that began at the same time as the N400. This frontally-distributed long-lasting effect was present during both first and second presentation of numerals, and took the form of a sustained negativity for numerals that participants self-rated as unfamiliar compared to numerals that they self-rated as familiar (see Figure 5.1). A first-pass interpretation of this effect might attribute it to the "FN400" commonly reported in studies of memory. In these reports, words recognized as old at test (familiar words) elicit less negative-going ERPs over frontal channels compared to words categorized as new at test (unfamiliar words). This is essentially a frontally distributed N400 repetition effect, which appears likely to reflect conceptual priming (Voss & Paller, 2009b; Voss & Federmeier, 2011). Additionally, the effect can be extended beyond the N400 timewindow (as in the present study), and it is less clear what that continued facilitation signifies (see, e.g., Wilding & Rugg, 1996; Woodruff, Hayama, & Rugg, 2006).

Two major issues with interpreting the present effect as a frontally-distributed N400 of the type reported in the memory literature is that the effect was not impacted by repetition (which would be expected of an N400 effect), and that we obtained it on first presentation (whereas in memory paradigms it is always reported on second presentation, during recall tests). That is, if the obtained effect truly reflected the kind of conceptual priming thought to be giving rise to the

FN400 effect in memory paradigms, then it should have been similarly affected by the priming induced by repetition. Because it was not, we suspect the difference between numerals rated familiar and unfamiliar must instead be caused by a different mechanism.

One possibility is that the frontal negativity for unfamiliar items is related to the family of sustained negativities that are often associated with differences in working memory load (with larger sustained negativities for larger loads) that are often reduced in age (e.g., Ruchkin, Canoune, Johnson, & Ritter, 2007; Chao & Knight, 1997). For example, sustained negativities are also a feature of difficult-to-parse sentences compared to less demanding sentences (King & Kutas, 1995). An additional domain that has shown similar ERP effects is studies of word ambiguity. Words that are ambiguous both in class (noun/verb) and semantic meaning (e.g., "duck", which could be utilized as a noun or verb, and neither interpretation conveys semantic overlap with the other) elicit larger frontally-distributed negativities as compared to words that are unambiguous in both their class and meaning (e.g., the noun "beer") (Federmeier, Segal, Lombrozo, & Kutas, 2000; Lee & Federmeier, 2006). Under this interpretation, then, numerals that are ultimately rated as unfamiliar might be read similarly to how ambiguous words are read (since both elicit frontally distributed negativities). It would also suggest that reading numerals with the aim of sorting them into meaningfulness categories involves a similar disambiguation process as is induced by word reading.

Clearly, more work is necessary to fully assess the possible relationship between word ambiguity effects, working memory processes, and the presently reported effects of familiarity judgment. Nevertheless, it is the case that one broadly applicable interpretation of our finding is that judging items that will ultimately be rated as unfamiliar is somehow more taxing. Whether this process is ultimately found to be related to working memory or not, perhaps the most interesting feature of the mechanisms involved in this effect is that they are independent of item repetition. That is, to the extent that this measure indicates some kind of processing effort, it is just as demanding to categorize unfamiliar items the second time as it was the first time they were encountered.

In contrast, there was a different effect that was both influenced by repetition and that distinguished between items rated familiar and items rated unfamiliar: the late positive complex (LPC). This effect was selectively enhanced (i.e., more positive) for second presentation of numerals rated as familiar (see Figure 5.2 for waveforms and Figure 5.3 for the topography of this effect). Items that were rated as unfamiliar only had an effect of repetition on the N400. One way to view the distinction between this pattern of N400 repetition effects (which were present regardless of how the item was rated) and LPC repetition effects (which were present only for numerals rated familiar) is that, whereas the N400 repetition effect measure primarily reflected *an attempt* to access numeral meaning, the LPC instead reflected actual retrieval success.

This perspective is consistent with the interpretation of LPC effects in memory paradigms, where ERPs are recorded during a test in which old (previously viewed) and new (previously unseen) items are classified by subjects. There, larger LPCs are typically associated with more successful recollection of old (repeated) items, especially when compared with successfully rejected new items or falsely rejected old items. Additionally, retrieval efforts which yield more explicit details about previously viewed items (e.g., which list they were seen on), and retrieval effects which result in higher-confidence judgments of an item being old, are associated with larger LPCs (see Voss & Paller, 2008, for review).

Our current finding is also strikingly similar to what has previously been observed when participants performed familiarity ratings for novel shapes: only novel shapes rated as familiar elicited LPC effects during a subsequent memory test (Voss et al., 2010). Of note is that there is a substantial difference between our paradigm and these memory paradigms. Whereas traditional memory paradigms clearly require more recollection and retrieval during the test (which is always second presentation) than during the encoding phase (first presentation), in our paradigm, first and second presentations both involved a memory test of sorts (to the extent that generating a rating of familiar/unfamiliar involved explicit long-term memory retrieval).

It is therefore interesting to see that this enhanced LPC selective to the second presentation of familiar items still persisted. This suggests that the recollection processes that occurred during first presentation for numerals that people rated as familiar somehow selectively facilitated the same retrieval processes at second presentation – either by augmenting the amount of information recollected or by rendering the judgment more confident (both of which are mechanisms that are associated with increased LPC size). Given that trials were only included if participants' judgments were consistent (e.g., if a participant only rated an item as familiar on second encounter, it was dropped from analysis), both interpretations (confidence/increased retrieval) are plausible – it seems unlikely that one would retrieve less information or be less confident when rating a numeral as familiar for the second time (especially in a relatively short amount of time). Most importantly, to our knowledge, this is the first piece of evidence linking a traditional measure of recollection (LPC) to not only to numerals, but to numerals that are specifically perceived by individuals as familiar due to personal (and not explicitly numerical) experience.

In sum, we found evidence that people are able to use numerals to access a different set of information than is typically studied (i.e., not specifically related to the numerosity or number value associated with the numeral as a symbol), manifesting across multiple effects when numerals were sorted by participants' subjective ratings of familiarity. There was a sustained frontal negativity larger for numerals rated unfamiliar, an LPC repetition effect larger for numerals rated familiar, and there was also an N400 repetition effect whose overall scalp distribution did not resemble the distribution previously observed during numeral reading in tasks wherein numeral familiarity was not being judged. However, some aspects of numeral processing were not sensitive to the act of making these familiarity judgments. Specifically, the size of the N400 repetition effect was unaffected by whether people rated the numerals as familiar or not, suggesting that the amount of meaning that is more effortlessly accessed from numerals is stable at an item level, even as a large-scale task manipulation might be able to impact the nature of the information being accessed (suggested by the distributional shift across experiments). These findings more broadly suggest that potentially non-numerical information related to our personal experience with Arabic numerals (e.g., apartment numbers, legally relevant age markers like 18 and 21, and numbers associated with sports figures) importantly shapes how we use numbers in everyday life, which is a relatively unexplored domain of numerical cognition.

Figures

Figure 5.1

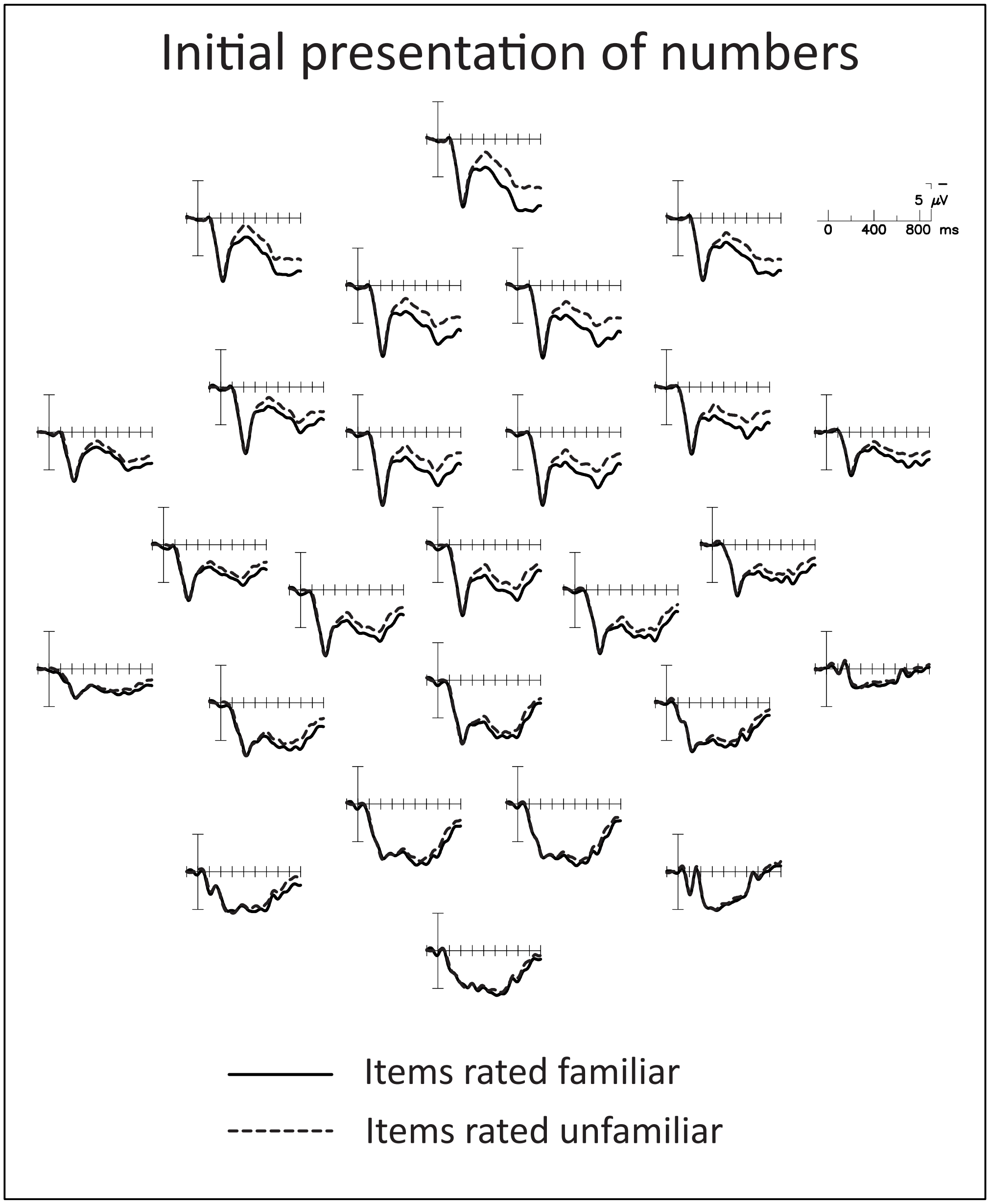

Figure 5.1 The average response to the first presentation of numerals, sorted by their subjective rating of familiarity. In frontal channels (top of figure), an effect of familiarity can be seen whereby numerals that were rated as unfamiliar across presentations elicited a long-lasting negativity compared to numerals that were rated familiar.

Figure 5.2

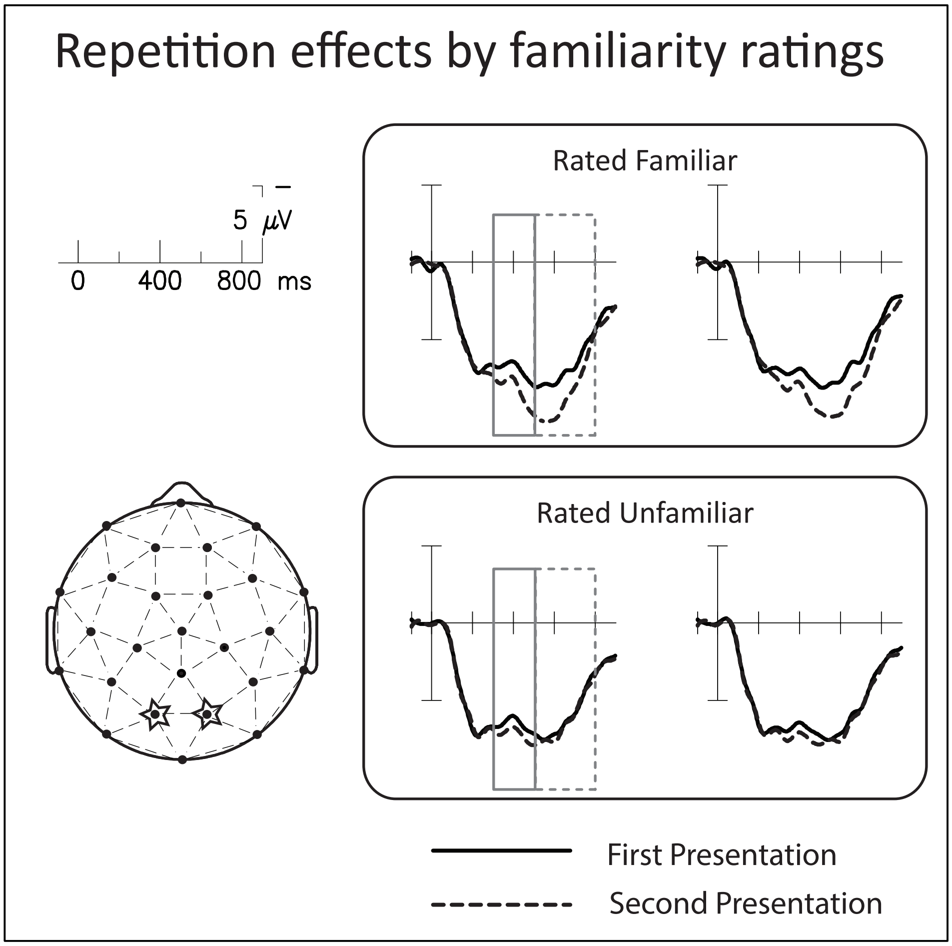

Figure 5.2 Left and right posterior channels are plotted (starred on the channel configuration plot at left). The solid box denotes the N400 timewindow, whereas the dashed box denotes the timewindow of the late positive complex (LPC). The top plot corresponds to repetition effects for numerals subjectively rated as familiar, reflected in the difference between second presentation compared to first presentation, and the bottom plot shows the same effects for numerals subjectively rated as unfamiliar. The LPC for second presentation is only for numerals rated as familiar, not numerals rated unfamiliar. The N400 repetition effect is small but present in all plots.

Figure 5.3

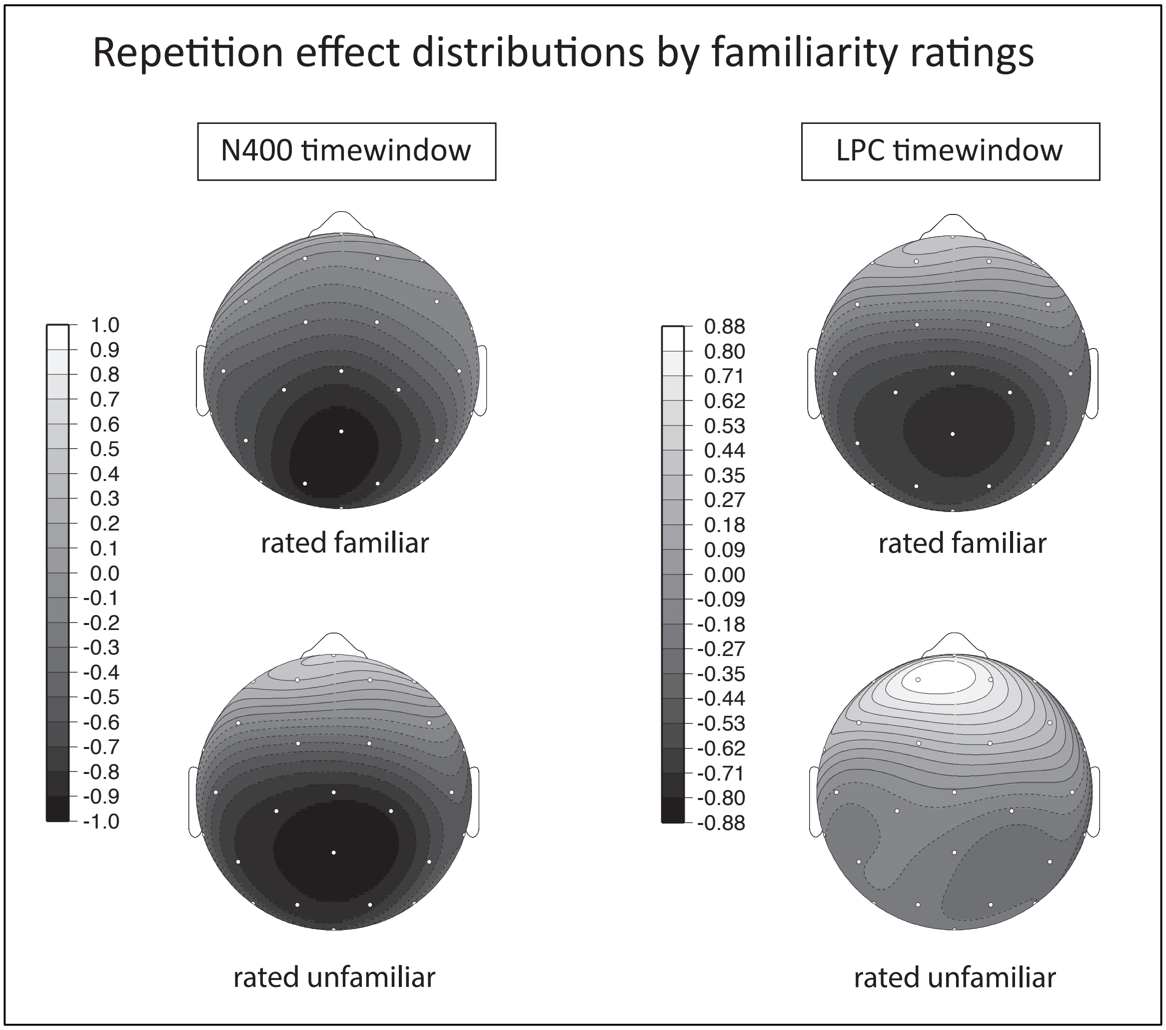

Figure 5.3 Topographical maps of the distribution of the effect of repetition, as measured by the subtraction of the first presentation of numerals from second presentation (viewed over the top of the head). Darker areas correspond to more substantial effects of repetition. On the left are repetition effects for the N400 timewindow (measured from 300 – 500ms post stimulus onset). On the right are repetition effects during the LPC timewindow (measured from 500 – 800ms post stimulus). It can be seen that there is an effect of repetition in N400 timewindow for numerals rated familiar and unfamiliar, but that in the LPC timewindow there is only a repetition effect for numerals rated familiar.

CHAPTER 6

Conclusion

Across four sets of experiments, we examined the flexible and task-specific utilization of multiple systems of memory during online reading of numerals in both neural and behavioral measures. The first two experiments focused on the processing of numerals in the context of their use in arithmetic problems, which depends on having built associations between numerals and arithmetic problems/operations in order to recognize whether a provided answer was correct or not. The third experiment looked at how people initially access the meaning of numerals across three tasks, measuring the benefit that prior access afforded incidental subsequent attempts of semantic access. Finally, a fourth experiment looked at how item-level experience might play a role in how people access and retrieve the meaning of numerals.

In Experiment 1, we tested how the hemispheres evaluated arithmetic problems by measuring brain responses (ERPs) to provided answers across levels of visual field (LH/RVF, RH/LVF, and Central VF). We expected that there should be differences across the hemispheres, particularly with respect to sensitivity to more detailed properties of answers than just whether they were correct or incorrect, given analogous research in the language domain (Federmeier & Kutas, 1999b). Indeed, we found that, although both hemispheres (individually and jointly) were sensitive to whether answers were correct or not in the same time frame (as measured on the P300 component, in the form of an amplitude difference), the hemispheres diverged in their responses to other aspects of answer type. In particular, only the RH/LVF distinguished between incorrect answers that were table-related to the correct answer and incorrect answers that were unrelated to the correct answer, in the form of a difference on a late positivity. We interpreted this as evidence that both hemispheres are involved in answer evaluation processes, and that the

RH/LVF may specifically perform additional deliberative processes beyond the initial categorization of an answer as matching what was expected or not.

However, what we could not know from this data was how these processing differences – which led to effects of correctness and relatedness effects on different componentry – might manifest in behavior. To address this question, in Experiment 2, we used the same materials from Experiment 1, but instead required participants to provide behavioral responses (respond as quickly as possible without sacrificing accuracy). In our analysis of these behavioral responses, we separately examined effects of the modal shift of the reaction time distribution from effects that altered the rightward skew of the underlying reaction time distribution. This revealed an interesting distinction between effects of correctness (which were selectively reflected on the modal parameter μ) and effects of relatedness (which were selectively reflected on the skewness parameter τ), which, in conjunction with the evidence from ERPs, support the idea that these two effects are not the result of identical processes. There were no effects of VF on any measure, despite the earlier finding that relatedness effects were only present in the RH/LVF. This suggests that the RH's special processing of these answer types (found in ERPs) is functionally relevant for behavior, and that, indeed, it is transferred to the other hemisphere (at least when motor responses from either hand are required by the task).

Together, Experiments 1 and 2 examined and identified evidence for explicit memory processing of arithmetic-related knowledge of numerals. What was still unclear from these experiments, however, was how semantic memory access (especially at a relatively automatic level) proceeded for numerals. Although this access presumably occurred in Experiment 1 as well, it was not possible to measure unambiguously because the ERP signal whose amplitude is typically taken as an index of initial semantic access (the N400) was obscured by the co-occurrence of the P300 effect. This left unclear whether initial semantic access for numerals was similar to that for words, and whether it could be facilitated by prior contextual support.

To this end, in Experiment 3, we examined the effect of implicit priming (as measured on N400 repetition effects) across multiple tasks that emphasized accessing different potential aspects or amounts of meaningfulness for numerals (visual appreciation of numeral content in the matching task, knowledge of arithmetic facts in the divisor task, and knowledge of everyday use of numerals in the quantifier task). There were three main findings: (1) a significant (but notably small) N400 repetition effect was obtained; (2) this effect's distribution across the scalp was atypical; and (3) the size of the effect of repetition was not affected by task. From these results, we inferred that, indeed, initial access of semantic meaning for numerals can be influenced by prior context (i.e., exposure at first presentation), and, further, that the benefit of prior access is not influenced by the amount/type of cognitive processing that occurred during initial presentation. The atypical scalp distribution of the effect that the scalp distribution of the effect suggested that the nature of the semantic information being accessed during numeral reading is potentially unique.

The small (but reliable) size of the effect, its atypical distribution, and its lack of interaction with task raised additional questions. If we changed the type of semantic access to be more similar to what occurs when we are trying to make deliberate connections with our broader semantic knowledge base (e.g., as we naturally do during word reading), could the distribution of the N400 change to be more canonical? Further, although the N400 effect was not affected by task manipulations, it was still possible that, at an item level, certain subsets of numerals elicited more consistent links to a broader sense of long-term semantic memory, which might emerge as larger N400 repetition effects compared to the subset of numerals that do not resonate as much with broader semantics.

Experiment 4 was conducted to address these possibilities. The same materials were used as in Experiment 3, with the only difference being that the task was now to classify items as personally meaningful or not. We obtained three additional findings in this study: (1) the scalp distribution of the N400 repetition effect for numerals shifted to what is typically observed in language studies; (2) the size of the N400 repetition effect was not influenced by item-level ratings of familiarity; and (3) there was an additional signature of recollection (LPC), specifically for repeated items that participants rated as familiar. We take the distributional shift across Experiments 3 and 4 as evidence that it is possible to alter the nature of the semantics that is being accessed when reading numerals, given that different processes would be necessary to manifest the observed effects. That the size of the N400 repetition effect itself was not impacted by self-rated familiarity with individual items implies that the amount of initial access of meaning for numerals is remarkably stable – consistent with the results of Experiment 3, where N400 repetition effect sizes were relatively stable across tasks. The overall smallness of these N400 effects across tasks and studies also is informative for interpretation of contextual effects in arithmetic – it seems likely that a considerable portion of the observed congruency effects arise from P300 activity rather than N400, as has been reported.

While initial access of meaning for items rated familiar and unfamiliar appears to have been the same, the enhanced LPC selectively for items rated familiar indicates that more confident recollection and explicit memory retrieval processes nevertheless occurred for these items at a later stage. Thus, when encountering numerals, multiple memory processes unfold – sometimes simultaneously. This is reminiscent of Experiment 1, where positivities revealed differences across types of numerals as a function of their contextual fit with preceding arithmetic contexts. There, numerals were similarly being explicitly judged and assessed.

There is an ongoing debate in the ERP literature about the nature and classification of these sorts of positivities (i.e., whether or not LPCs and other late positivities are members of a broad family of P300s) (Coulson, King, & Kutas, 1998; Van Petten & Luka, 2012). While our results cannot directly speak to this debate, we have shown that it is critical to be conscious of whether or not explicit decision-making-related componentry, which often manifest as positivities, might emerge in a study. This is particularly important if what is of greatest interest are relatively more automatic semantic memory access processes (as measured best by the N400).

In general, using a variety of tasks and paradigms that deliberately vary on how much they rely on implicit/automatic versus explicit/controlled memory access has helped to better delineate the possible meaning people can extract from, or appreciate in, numerals. This is particularly true of the experiments presented here because the measurements were taken largely on well-defined components from across the broader ERP literature (N400, P300, LPC), whose links to automatic and controlled processing mechanisms are well documented. Even so, some unexpected results we obtained and reported were on components that are not as well understood (e.g., a frontal negativity for numerals rated unfamiliar). Using numerals as stimuli can thus inform how generalizable effects – and interpretations of those effects – are across different domains (e.g., maybe frontal negativities that have been attributed to word ambiguity reflect other kinds of uncertainty in stimulus evaluation as well).

In sum, our work has made important contributions to differentiating between how we initially, effortless access the meaning of numerals and how we can later go on to utilize deliberately retrieved memories for them when rendering explicit judgments.

REFERENCES

Anderson, S. W., Damasio, A. R., & Damasio, H. (1990). Troubled letters but not numbers: domain specific cognitive impairments following focal damage in frontal cortex. *Brain, 113*(3), 749-766.

Andres, M., Pelgrims, B., Michaux, N., Olivier, E., & Pesenti, M. (2011). Role of distinct parietal areas in arithmetic: An fMRI-guided TMS study. *Neuroimage, 54*(4), 3048-3056.

Andres, M., Seron, X., & Olivier, E. (2005). Hemispheric lateralization of number comparison. *Cognitive Brain Research, 25*(1), 283-290.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1), 75-106.

Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition, 1*(1), 3-34.

Ashcraft, M. H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4*(5), 527-538.

Ashcraft, M. H., & Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in grades 1-6. *Journal for Research in Mathematics Education, 26*(5), 396-421.

Ashcraft, M. H., Fierman, B. A., & Bartolotta, R. (1984). The production and verification tasks in mental addition: An empirical comparison. *Developmental Review, 4*(2), 157-170.

Ashcraft, M. H., Yamashita, T. S., & Aram, D. M. (1992). Mathematics performance in left and right brain-lesioned children and adolescents. *Brain and Cognition, 19*(2), 208-252.

Babkoff, H., & Faust, M. (1988). Lexical decision and visual hemifield: An examination of the RT—Accuracy relationship. *Neuropsychologia, 26*(5), 711-725.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*(1), 32-55.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry the power of response time distributional analyses. *Current Directions in Psychological Science, 20*(3), 160-166.

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language, 59*(4), 495-523.

Banich, M. T. (2002). The divided visual field technique in laterality and interhemispheric integration. In K. Hugdahl (Ed.), *Experimental methods in neuropsychology* (pp. 47-64) Boston, MA: Klewer.

Barnett, K. J., & Corballis, M. C. (2005). Speeded right-to-left information transfer: The result of speeded transmission in right-hemisphere axons? *Neuroscience Letters, 380*(1), 88-92.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*(2), 201-212.

Barrouillet, P., Fayol, M., & Lathulière, E. (1997). Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning difficulties. *International Journal of Behavioral Development, 21*(2), 253-275.

Bentin, S., & Peled, B. (1990). The contribution of task-related factors to ERP repetition effects at short and long lags. *Memory & Cognition, 18*(4), 359-366.

Besson, M., & Macar, F. (1987). An Event-Related potential analysis of incongruity in music and other Non-Linguistic contexts. *Psychophysiology, 24*(1), 14-25.

Binder, J. R., Swanson, S. J., Hammeke, T. A., Morris, G. L., Mueller, W. M., Fischer, M., et al. (1996). Determination of language dominance using functional MRI: A comparison with the Wada test. *Neurology, 46*(4), 978-984.

Bishop, D. V. (2013). Cerebral asymmetry and language development: Cause, correlate, or consequence? *Science (New York, NY), 340*(6138), 1230531.

Blankenberger, S., & Vorberg, D. (1997). The single-format assumption in arithmetic fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 721-738.

Brannon, E. M., & Roitman, J. D. (2003). Nonverbal representations of time and number in animals and human infants. In W. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 143-182). Boca Raton, FL: CRC Press.

Broca, P. (1865). Sur le siège de la faculté du langage articulé. *Bulletins De La Société d'Anthropologie De Paris, 6*(1), 377-393.

Brunia, C. H., van Boxtel, G. M., & Böcker, K. B. (2012). Negative slow waves as indices of anticipation: The bereitschaftspotential, the contingent negative variation, and the stimulus-preceding negativity. In E. S. Kappenman, & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 189-207). New York, NY: Oxford University Press.

Brysbaert, M. (2005). Number recognition in different formats. In J. I. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 23-42). New York, NY: Psychology Press.

Campbell, J. I. (1987a). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 109-123.

Campbell, J. I. (1987b). Production, verification, and priming of multiplication facts. *Memory & Cognition, 15*(4), 349-364.

Campbell, J. I., & Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 39*(2), 338-366.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia, 39*(11), 1224-1239.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2012). Numeracy skills in patients with degenerative disorders and focal brain lesions: A neuropsychological investigation. *Neuropsychology, 26*(1), 1-19.

Cappelletti, M., & Cipolotti, L. (2010). The neuropsychology of acquired calculation disorders. In J. M. Gurd, U. Kischka, & J. C. Marshall (Eds.), *Handbook of Clinical Neuropsychology (*pp. 401-417). New York, NY: Oxford University Press.

Cappelletti, M., Kopelman, M. D., Morton, J., & Butterworth, B. (2005). Dissociations in numerical abilities revealed by progressive cognitive decline in a patient with semantic dementia. *Cognitive Neuropsychology, 22*(7), 771-793.

Castro, A., Sumich, A., Premkumar, P., & Jones, G. (2014). How do incorrect results change the processing of arithmetic information? evidence from a divided visual field experiment. *Laterality: Asymmetries of Body, Brain and Cognition, 19*(3), 340-353.

Chao, L. L., & Knight, R. T. (1997). Prefrontal deficits in attention and inhibitory control with aging. *Cerebral Cortex (New York, N.Y.: 1991), 7*(1), 63-69.

Chochon, F., Cohen, L., Van De Moortele, P., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience, 11*(6), 617-630.

Christman, S., Kitterle, F. L., & Hellige, J. (1991). Hemispheric asymmetry in the processing of absolute versus relative spatial frequency. *Brain and Cognition, 16*(1), 62-73.

Cipolotti, L. (1995). Multiple routes for reading words, why not numbers? evidence from a case of arabic numeral dyslexia. *Cognitive Neuropsychology, 12*(3), 313-342.

Cipolotti, L., Warrington, E. K., & Butterworth, B. (1995). Selective impairment in manipulating Arabic numerals. *Cortex, 31*(1), 73-86.

Cohen, L., Dehaene, S., Chochon, F., Lehericy, S., & Naccache, L. (2000). Language and calculation within the parietal lobe: A combined cognitive, anatomical and fMRI study. *Neuropsychologia, 38*(10), 1426-1440.

Colvin, M. K., Funnell, M. G., & Gazzaniga, M. S. (2005). Numerical processing in the two hemispheres: Studies of a split-brain patient. *Brain and Cognition, 57*(1), 43-52.

Corina, D., Grosvald, M., & Lachaud, C. (2011). Perceptual invariance or orientation specificity in american sign language? evidence from repetition priming for signs and gestures. *Language and Cognitive Processes, 26*(8), 1102-1135.

Coulson, S. (1998). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13*(6), 653-672.

Coulson, S., Federmeier, K. D., Van Petten, C., & Kutas, M. (2005). Right hemisphere sensitivity to word-and sentence-level context: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 129-147.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Crutch, S. J., & Warrington, E. K. (2002). Preserved calculation skills in a case of semantic dementia. *Cortex, 38*(3), 389-399.

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research, 15*(2), 191-205.

Dale, A. M. (1994). *Source localization and spatial discriminant analysis of event-related potentials: linear approaches.* Unpublished Dissertation, University of California, San Diego, La Jolla, CA.

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21*(10), 1882-1892.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychological Research, 62*(1), 72-77.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122*(3), 371-396.

Dehaene, S., & Changeux, J. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*(4), 390-407.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition, 1*(1), 83-120.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21*(8), 355-361.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 626-641.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science (New York, N.Y.), 284*(5416), 970-974.

Delazer, M., Girelli, L., Semenza, C., & Denes, G. (1999). Numerical skills and aphasia. *Journal of the International Neuropsychological Society, 5*(3), 213-221.

Dickson, D. S., & Federmeier, K. D. (2014). Hemispheric differences in orthographic and semantic processing as revealed by event-related potentials. *Neuropsychologia, 64*, 230-239.

Diesfeldt, H. F. (1993). Progressive decline of semantic memory with preservation of number processing and calculation. *Behavioural Neurology, 6*(4), 239-242.

Domahs, F., Delazer, M., & Nuerk, H. (2006). What makes multiplication facts difficult: Problem size or neighborhood consistency? *Experimental Psychology, 53*(4), 275-282.

Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., et al. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behavioral and Brain Functions, 3*(1), 66.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating. *Behavioral and Brain Sciences, 11*(3), 357-427.

Duzel, E., Vargha-Khadem, F., Heinze, H. J., & Mishkin, M. (2001). Brain activity evidence for recognition without recollection after early hippocampal damage. *Proceedings of the National Academy of Sciences of the United States of America, 98*(14), 8101-8106.

Duzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H. J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences of the United States of America, 94*(11), 5973-5978.

Eddy, M., Schmid, A., & Holcomb, P. J. (2006). Masked repetition priming and event-related brain potentials: A new approach for tracking the time-course of object perception. *Psychophysiology, 43*(6), 564-568.

Fayol, M., & Seron, X. (2005). About numerical representations: Insights from neuropsychological, experimental, and developmental studies. In J. I. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 3-22). New York, NY: Psychology Press.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999a). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

Federmeier, K. D., & Kutas, M. (1999b). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research, 8*(3), 373-392.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 202-224.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia, 40*(7), 730-747.

Federmeier, K. D., Kutas, M., & Dickson, D. S. (2015). A common neural progression to meaning in about a third of a second. In G. S. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 557-568). Holland: Elsevier.

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. *Memory & Cognition, 33*(5), 871-886.

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133-146.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123*(12), 2552-2566.

Fias, W., Brysbaert, M., Geypens, F., & D'ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2*(1), 95-110.

Fias, M. & Fischer, M. H. (2005). Spatial representation of numbers. In J. I. Campbell (Ed.), *The Handbook of Mathematical Cognition* (pp. 43-54). New York, NY: Psychology Press.

Fias, W., van Dijck, J., & Gevers, W. (2011). How number is associated with space?: The role of working memory. In S. Dehaene & E. M. Brannon (Eds.), *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought* (pp. 133-148). Amsterdam: Academic Press – Elsevier.

Fischer-Baum, S., Dickson, D. S., & Federmeier, K. D. (2014). Frequency and regularity effects in reading are task dependent: Evidence from ERPs. *Language, Cognition and Neuroscience, 29*(10), 1342-1355.

Funnell, M. G., Colvin, M. K., & Gazzaniga, M. S. (2007). The calculating hemispheres: Studies of a split-brain patient. *Neuropsychologia, 45*(10), 2378-2386.

Gaber, D., & Schlimm, D. (2015). Basic mathematical cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(4), 355-369.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1), 43-74.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*(2), 123-144.

Ganis, G., Kutas, M., & Sereno, M. E. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Cognitive Neuroscience, Journal of, 8*(2), 89-106.

Ganor-Stern, D., Tzelgov, J., & Ellenbogen, R. (2007). Automaticity and two-digit numbers. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 483-496.

García-Orza, J., Perea, M., Mallouh, R. A., & Carreiras, M. (2012). Physical similarity (and not quantity representation) drives perceptual comparison of numbers: Evidence from two Indian notations. *Psychonomic Bulletin & Review, 19*(2), 294-300.

Gazzaniga, M. S., & Smylie, C. S. (1984). Dissociation of language and cognition. A psychological profile of two disconnected right hemispheres. *Brain, 107*(1), 145-153.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9*(1), 6-10.

Geschwind, N., & Levitsky, W. (1968). Human brain: Left-right asymmetries in temporal speech region. *Science (New York, N.Y.), 161*(3837), 186-187.

Geyer, A., Holcomb, P., Kuperberg, G., & Perlmutter, N. (2006, April). Plausibility and sentence comprehension. An ERP study. Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

Goldfarb, L., Henik, A., Rubinsten, O., Bloch-David, Y., & Gertner, L. (2011). The numerical distance effect is task dependent. *Memory & Cognition, 39*(8), 1508-1517.

Gross-Tsur, V., Manor, O., & Shalev, R. S. (1996). Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine & Child Neurology, 38*(1), 25-33.

Hamann, M. S., & Ashcraft, M. H. (1986). Textbook presentations of the basic addition facts. *Cognition and Instruction, 3*(3), 173-202.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition, 10*(4), 389-395.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24*(2), 259-276.

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience, 14*(6), 938-950.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience, 18*(10), 1631-1643.

Holcomb, P. J., & Grainger, J. (2007). Exploring the temporal dynamics of visual word recognition in the masked repetition priming paradigm using event-related potentials. *Brain Research, 1180*, 39-58.

Holloway, I. D., & Ansari, D. (2010). Developmental specialization in the right intraparietal sulcus for the abstract representation of numerical magnitude. *Journal of Cognitive Neuroscience, 22*(11), 2627-2637.

Jackson, J. D., Balota, D. A., Duchek, J. M., & Head, D. (2012). White matter integrity and reaction time intraindividual variability in healthy aging and early-stage Alzheimer disease. *Neuropsychologia, 50*(3), 357-366.

Jackson, M., & Warrington, E. K. (1986). Arithmetic skills in patients with unilateral cerebral lesions. *Cortex, 22*(4), 611-620.

Jasinski, E. C., & Coch, D. (2012). ERPs across arithmetic operations in a delayed answer verification task. *Psychophysiology, 49*(7), 943-958.

Johnson, R. (1986). For distinguished early career contribution to psychophysiology: Award address, 1985. *Psychophysiology, 23*(4), 367-384.

Jordan, T. R., Patching, G. R., & Milner, A. D. (2000). Lateralized word recognition: Assessing the role of hemispheric specialization, modes of lexical access, and perceptual asymmetry. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 1192-1208.

Jordan, T. R., Patching, G. R., & Thomas, S. M. (2003). Assessing the role of hemispheric specialisation, serial-position processing, and retinal eccentricity in lateralised word recognition. *Cognitive Neuropsychology, 20*(1), 49-71.

Joseph B. Hellige. (1993). *Hemispheric asymmetry: What's right and what's left*. Cambridge, MA: Harvard University Press.

Joyce, C. A., Paller, K. A., Schwartz, T. J., & Kutas, M. (1999). An electrophysiological analysis of modality-specific aspects of word repetition. *Psychophysiology, 36*(5), 655-665.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences, 9*(11), 512-518.

Kadosh, R. C., Kadosh, K. C., Schuhmann, T., Kaas, A., Goebel, R., Henik, A., et al. (2007). Virtual dyscalculia induced by parietal-lobe TMS impairs automatic magnitude processing. *Current Biology, 17*(8), 689-693.

Kadosh, R. C., Tzelgov, J., & Henik, A. (2008). A synesthetic walk on the mental number line: The size effect. *Cognition, 106*(1), 548-557.

Kadosh, R. C., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences, 32*(3-4), 313-328.

Kandhadai, P., & Federmeier, K. D. (2007). Multiple priming of lexically ambiguous and unambiguous targets in the cerebral hemispheres: The coarse coding hypothesis revisited. *Brain Research, 1153*, 144-157.

Kandhadai, P., & Federmeier, K. D. (2008). Summing it up: Semantic activation processes in the two hemispheres as revealed by event-related potentials. *Brain Research, 1233*, 146-159.

Kandhadai, P., & Federmeier, K. D. (2010). Automatic and controlled aspects of lexical associative processing in the two cerebral hemispheres. *Psychophysiology, 47*(4), 774-785.

Kandhadai, P., & Federmeier, K. D. (2010). Hemispheric differences in the recruitment of semantic processing mechanisms. *Neuropsychologia, 48*(13), 3772-3781.

Katz, A. N. (1980). Cognitive arithmetic: Evidence for right hemispheric mediation in an elementary component stage. *Quarterly Journal of Experimental Psychology, 32*(1), 69-84.

King, J. W., & Kutas, M. (1995). Who did what and when? using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*(3), 376-395.

Klein, R., & McInnes, J. (1988). Visual field differences in the processing of numerical stimuli. *Brain and Cognition, 7*(3), 247-256.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302-307.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science (New York, N.Y.), 197*(4305), 792-795.

Laszlo, S., & Armstrong, B. C. (2014). PSPs and ERPs: Applying the dynamics of post-synaptic potentials to individual units in simulation of temporally extended event-related potential reading data. *Brain and Language, 132*, 22-27.

Laszlo, S., & Federmeier, K. D. (2007). Better the DVL you know: Acronyms reveal the contribution of familiarity to single-word reading. *Psychological Science, 18*(2), 122-126.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61*(3), 326-338.

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48*(2), 176-186.

Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: Evaluating the physiological time course of visual word recognition with regression analysis of single-item event-related potentials. *Language, Cognition and Neuroscience, 29*(5), 642-661.

Laszlo, S., Stites, M., & Federmeier, K. D. (2012). Won't get fooled again: An event-related potential study of task and repetition effects on the semantic processing of items without semantics. *Language and Cognitive Processes, 27*(2), 257-274.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920-933.

Lee, C., & Federmeier, K. D. (2006). To mind the mind: An event-related potential study of word class and semantic ambiguity. *Brain Research, 1081*(1), 191-202.

Lee, C., & Federmeier, K. D. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language, 104*(2), 145-158.

LeFevre, J., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General, 125*(3), 284-306.

LeFevre, J., Smith-Chant, B. L., Hiscock, K., Daley, K. E., & Morris, J. (2003). Young adults' strategic choices in simple arithmetic: Implications for the development of mathematical representations. In E. J. Baroody, & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructive adaptive expertise* (pp. 203-228). Mahwah, NJ: Erlbaum.

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? the example of parity and five effects on product verification. *Memory & Cognition, 27*(2), 364-382.

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behavioral and Brain Functions, 3*(1), 1.

Lochy, A., Seron, X., Delazer, M., & Butterworth, B. (2000). The odd-even effect in multiplication: Parity rule or familiarity with even numbers? *Memory & Cognition, 28*(3), 358-365.

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General, 141*(4), 635-641.

Martin, M. (1979). Hemispheric specialization for local and global processing. *Neuropsychologia, 17*(1), 33-40.

Marzi, C. A., Bisiacchi, P., & Nicoletti, R. (1991). Is interhemispheric transfer of visuomotor information asymmetric? evidence from a meta-analysis. *Neuropsychologia, 29*(12), 1163-1177.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science (New York, N.Y.), 211*(4477), 77-80.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*(1), 107-157.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*(2), 171-196.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*(1), 53-65.

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to "d'oh!": Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 525-549.

Nagy, M. E., & Rugg, M. D. (1989). Modulation of Event-Related potentials by word repetition: The effects of Inter-Item lag. *Psychophysiology, 26*(4), 431-436.

Niedeggen, M., & Rösler, F. (1999). N400 effects reflect activation spread during retrieval of arithmetic facts. *Psychological Science, 10*(3), 271-276.

Niedeggen, M., Rösler, F., & Jost, K. (1999). Processing of incongruous mental calculation problems: Evidence for an arithmetic N400 effect. *Psychophysiology, 36*(03), 307-324.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science (New York, N.Y.), 297*(5587), 1708-1711.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15-22.

Nuerk, H., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? putting the tens and units back in different bins. *Cognition, 82*(1), B25-B33.

Núñez, R. E. (2011). No innate number line in the human brain. *Journal of Cross-Cultural Psychology, 42*(4), 651-668.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia. electrophysiological measures of impaired and spared memory. *Brain, 123*(9), 1948-1963.

Paller, K. A., Kutas, M., & McIsaac, H. K. (1995). Monitoring conscious recollection via the electrical activity of the brain. *Psychological Science, 6*(2), 107-111.

Paller, K. A., McCarthy, G., & Wood, C. C. (1992). Event-Related potentials elicited by deviant endings to melodies. *Psychophysiology, 29*(2), 202-206.

Pavese, A., & Umiltà, C. (1998). Symbolic distance between numerosity and identity modulates stroop interference. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1535-1545.

Payne, B. R., Lee, C., & Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: Evidence from single-word event-related brain potentials. *Psychophysiology, 52*(11), 1456-1469.

Payne, B. R., & Stine-Morrow, E. A. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-Gaussian distributional analyses of reading time. *Psychology and Aging, 29*(2), 213-228.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage, 62*(2), 816-847.

Ragot, R., & Renault, B. (1981). P300, as a function of S—R compatibility and motor programming. *Biological Psychology, 13*, 289-294.

Ranganath, C., & Paller, K. A. (1999). Frontal brain potentials during recognition are modulated by requirements to retrieve perceptual detail. *Neuron, 22*(3), 605-613.

Ratcliff, R., & Hacker, M. J. (1981). Speed and accuracy of same and different responses in perceptual matching. *Perception & Psychophysics, 30*(3), 303-307.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347-356.

Reingold, E. M., Sheridan, H., & Reichle, E. D. (2015). 18 direct lexical and nonlexical control of fixation duration in reading. In A. Pollatsek & R. Treiman (Eds), *The Oxford Handbook of Reading* (pp. 261-276). New York, NY: Oxford University Press.

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit arabic numerals address the same semantic number line. *Cognition, 72*(2), 191-201.

Ritter, W., Simson, R., & Vaughan, H. G. (1972). Association cortex potentials and reaction time in auditory discrimination. *Electroencephalography and Clinical Neurophysiology, 33*(6), 547-555.

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., & Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia, 49*(9), 2592-2608.

Rosselli, M., & Ardila, A. (1989). Calculation deficits in patients with right and left hemisphere damage. *Neuropsychologia, 27*(5), 607-617.

Rossor, M. N., Warrington, E. K., & Cipolotti, L. (1995). The isolation of calculation skills. *Journal of Neurology, 242*(2), 78-81.

Ruchkin, D. S., Canoune, H. L., Johnson, R., & Ritter, W. (1995). Working memory and preparation elicit different patterns of slow wave event-related brain potentials. *Psychophysiology, 32*(4), 399-410.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition, 18*(4), 367-379.

Rugg, M. D., & Doyle, M. C. (1994). Event-related potentials and stimulus repetition in direct and indirect tests of memory. In H. -J. Heinze, T. F. Münte, & G. R. Mangun (Eds.), *Cognitive Electrophysiology* (pp. 124-148). Cambridge, MA: Birkhäuser Boston.

Rugg, M. D., Doyle, M. C., & Melan, C. (1993). An event-related potential study of the effects of within-and across-modality word repetition. *Language and Cognitive Processes, 8*(4), 357-377.

Rugg, M. D., Furda, J., & Lorist, M. (1988). The effects of task on the modulation of event-related potentials by word repetition. *Psychophysiology, 25*(1), 55-63.

Rugg, M. D., & Nagy, M. E. (1987). Lexical contribution to nonword-repetition effects: Evidence from event-related potentials. *Memory & Cognition, 15*(6), 473-481.

Rugg, M. D., & Nagy, M. E. (1989). Event-related potentials and recognition memory for words. *Electroencephalography and Clinical Neurophysiology, 72*(5), 395-406.

Sandrini, M., Rossini, P. M., & Miniussi, C. (2004). The differential involvement of inferior parietal lobule in number comparison: A rTMS study. *Neuropsychologia, 42*(14), 1902-1909.

Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition, 108*(1), 263-270.

Schacter, D. L. (1987). Implicit expressions of memory in organic amnesia: Learning of new facts and associations. *Human Neurobiology, 6*(2), 107-118.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience, 15*(1), 111-135.

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience, 19*(5), 734-749.

Schendan, H. E., & Lucia, L. C. (2010). Object-sensitive activity reflects earlier perceptual and later cognitive processing of visual objects between 95 and 500ms. *Brain Research, 1329*, 124-141.

Schirmer, A., Soh, Y. H., Penney, T. B., & Wyse, L. (2011). Perceptual and conceptual priming of environmental sounds. *Journal of Cognitive Neuroscience, 23*(11), 3241-3253.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*(3), 414-429.

Schweinberger, S. R., Huddy, V., & Burton, A. M. (2004). N250r: A face-selective brain response to stimulus repetitions. *Neuroreport, 15*(9), 1501-1505.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research, 14*(3), 398-409.

Sergent, J., & Hellige, J. B. (1986). Role of input factors in visual-field asymmetries. *Brain and Cognition, 5*(2), 174-199.

Simmons, F. R., & Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia, 14*(2), 77-94.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real–world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

Smith, M. E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgments. *Journal of Cognitive Neuroscience, 5*(1), 1-13.

Smith, M. E., & Halgren, E. (1987). Event-related potentials during lexical decision: Effects of repetition, word frequency, pronounceability, and concreteness. *Electroencephalography and Clinical Neurophysiology, Supplement, 40*, 417-421.

Spelke, E., & Dehaene, S. (1999). Biological foundations of numerical thinking. *Trends in Cognitive Sciences, 3*(10), 365-366.

Staub, A., & Benatar, A. (2013). Individual differences in fixation duration distributions in reading. *Psychonomic Bulletin & Review, 20*(6), 1304-1311.

Stazyk, E. H., Ashcraft, M. H., & Hamann, M. S. (1982). A network approach to mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(4), 320-335.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382-392.

Tse, C., Balota, D. A., Yap, M. J., Duchek, J. M., & McCabe, D. P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology, 24*(3), 300-315.

Umiltà, C., Priftis, K., & Zorzi, M. (2009). The spatial representation of numbers: Evidence from neglect and pseudoneglect. *Experimental Brain Research, 192*(3), 561-569.

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241-255.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequencyinevent-related brainpotentials. *Memory & Cognition, 18*(4), 380-393.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Van Petten, C., & Senkfor, A. J. (1996). Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. *Psychophysiology, 33*(5), 491-506.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7*(3), 424-465.

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition, 33*(1), 1-16.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1656-1674.

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology, 48*(4), 532-546.

Voss, J. L., & Paller, K. A. (2006). Fluent conceptual processing and explicit memory for faces are electrophysiologically distinct. *The Journal of Neuroscience, 26*(3), 926-933.

Voss, J., & Paller, K. (2008). Neural substrates of Remembering – electroencephalographic studies. In J Bryne (Ed.), *Learning and Memory: A comprehensive reference* (Vol. 3, pp. 79-97). Oxford: Elsevier.

Voss, J. L., & Paller, K. A. (2009a). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience, 12*(3), 349-355.

Voss, J. L., & Paller, K. A. (2009b). Remembering and knowing: Electrophysiological distinctions at encoding but not retrieval. *Neuroimage, 46*(1), 280-289.

Voss, J. L., Schendan, H. E., & Paller, K. A. (2010). Finding meaning in novel geometric shapes influences electrophysiological correlates of repetition and dissociates perceptual and conceptual priming. *Neuroimage, 49*(3), 2879-2889.

Wada, J., & Rasmussen, T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: Experimental and clinical observations. *Journal of Neurosurgery, 17*(2), 266-282.

Warrington, E. K. (1982). The fractionation of arithmetical skills: A single case study. *The Quarterly Journal of Experimental Psychology, 34*(1), 31-51.

Warrington, E. K., & Weiskrantz, L. (1982). Amnesia: A disconnection syndrome? *Neuropsychologia, 20*(3), 233-248.

Winkelman, J. H., & Schmidt, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology, 102*(4), 734-736.

Wlotko, E. W., & Federmeier, K. D. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia, 45*(13), 3001-3014.

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Research, 1100*, 125-135.

Woroch, B., & Gonsalves, B. D. (2010). Event-related potential correlates of item and source memory strength. *Brain Research, 1317*, 180-191.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654-667.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General, 115*(2), 118-130.

Zbrodoff, N. J., & Logan, G. D. (2005). What everyone finds: The problem-size effect. In J. I. Campbell (Ed.), *Handbook of Mathematical Cognition* (pp. 331-345). New York, NY: Psychology Press.

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 778-783). Mahwah, NJ: Erlbaum.