Research report

The influence of increased working memory load on semantic neural systems: a high-resolution event-related brain potential study

Ryan C.N. D’Arcya,b,c,*, Elisabet Serviceb,d, John F. Connollyb,c,f,g, Colin S. Hawcoh

*Corresponding author. National Research Council, Institute for Biodiagnostics (Atlantic), National Research Council, Halifax, NS, Canada
bDepartment of Psychology, Dalhousie University, Halifax, NS, Canada
cDepartment of Radiology, Dalhousie University, Halifax, NS, Canada
dDepartment of Psychology, University of Helsinki, Helsinki, Finland
eDepartment of Psychiatry, Dalhousie University, Halifax, NS, Canada
fPediatrics (Neurology), IWK Health Centre, Dalhousie University, Halifax, NS, Canada
gDepartment of Medicine (Neurology), Dalhousie University, Halifax, NS, Canada
hMontreal Neurological Institute, McGill University, Montreal, PQ, Canada

Accepted 17 August 2004
Available online 13 September 2004

Abstract

The effects of working memory (WM) on the semantic N400 response were studied using high-resolution event-related brain potentials (ERPs). Participants were presented with semantically related sentence pairs and the terminal word congruence was varied in the second sentence. WM load was varied for the sentence pairs using a modified fan procedure [J.R. Anderson, Retrieval of propositional information from long-term memory. Cogn. Psychol., 6 (1974) 451–474; J.R. Anderson, A spreading activation theory of memory. J. Verbal Learn. Verbal Behav., 22 (1983) 261–295]. ERPs were recorded to the onset of the terminal word (Congruent and Incongruent) for sentence pairs at two levels of WM load (WML1 and WML2). Behavioral data analysis revealed that reaction times (RTs) increased as a function of WM load (i.e., the fan effect). Semantically incongruent words elicited an N400 response. Increased WM load reduced the congruency effect and, importantly, resulted in a significant delay in N400 peak latency (~50 ms). Moreover, the latency delay was correlated with a neuropsychological measure of individual WM capacity. WM load effects on the N400 were interpreted as a result of competing activation in WM, both modulating semantic expectancies and delaying semantic integration. Exploratory source analysis revealed activation in occipital, temporal, and parietal regions. Consistent with prior reports on the N400 and semantic processing, equivalent current dipoles were modeled in the left perisylvian region. Increased WM load led to novel source activation in the left inferior parietal region as well as increased activation levels in anterior temporal sources. The source modeling results were in agreement with the functional roles typically ascribed to these areas and confirmed that the scalp-recorded WM load effects on the N400 were present within the intracranial generators.

© 2004 Elsevier B.V. All rights reserved.

Theme: Neural basis of behavior
Topic: Learning and memory: physiology

Keywords: ERP; N400; Cortical current source estimation; Working memory; Fan effect; Semantic processing; Reading; Language

1. Introduction

Numerous studies point to a relation between individual working memory (WM) capacity, as measured with tasks combining a processing and a storage component, and language comprehension [19,20,27]. However, the manner in which WM and semantic comprehension...
interact remains to be determined. Behavioral studies have developed theories concerning the relationship between the two cognitive processes, with two main hypotheses being contrasted. According to one, skilled comprehension requires little WM capacity and results in good memory performance in tests that combine language processing with storage of single words, so-called complex span tests. Thus, a good WM score in complex span tests is a result of good performance in its processing (often language comprehension) component [21]. Other research suggests that true WM capacity, or capacity of executive attention in the face of interference, drives the relationship, being needed in both comprehension and the performance of complex span tasks [27,28]. A third view, not incompatible with the latter hypothesis, is that complex span tasks involve storage tasks that are interrupted by other activity and that the capacity to cope with this interference is related to performance in other cognitive tasks [14]. Recent advances in neuroimaging techniques have allowed for the neural correlates of language comprehension processes to be identified. This makes it possible to add neuroimaging techniques to the arsenal of behavioral methods in exploring the complex relationship between semantic processing and WM.

In the current investigation, we used a WM sensitive retrieval paradigm, showing the so-called fan effect [2,3]. This effect has been explained as resulting from competing information in long-term memory (LTM) giving rise to an extra WM load [17]. The interaction of this load with sentence comprehension was studied by recording the time behavior, and localizing the generators of an event-related brain potential (ERP): the semantic N400 response [48,49].

1.1. WM, comprehension, and the behavioral fan effect

Working memory is thought to be a system involved in maintaining newly acquired and reactivating stored verbal and nonverbal information in order to make it available for processing [4–6]. Multiple theories about WM exist (e.g., [3,18,26,29,31]), all of which address the relationship with complex cognitive processes [53]. Different theoretical perspectives on WM present alternative views concerning WM mechanisms, control, non-unitary nature, limitations, interaction with cognition, relationship to LTM, role of attention, and biological basis [53]. Despite different terminologies and architectures, much of the current work can be couched in terms of the Baddeley and Hitch framework [6]. A central part of this research has focused on the relation between WM and complex cognitive processes, especially semantic comprehension.

Behavioral evidence has demonstrated the involvement of WM in comprehension abilities. Daneman and Carpenter [19] developed a method for estimating individual WM capacity. The Reading Span task combined concurrent processing of sentences with storage of words. The original study, as well as numerous subsequent studies [20], found that individual differences in reading comprehension abilities correlate with WM capacity (as measured by the Reading Span and similar tasks). King and Just [45] used object-relative and subject-relative sentences to show that individual differences in syntactic processing were also related to differences in WM capacity. In an attempt to better understand the relationship between WM and comprehension, Engle et al. [13,17] examined the relationship between individual WM capacity and reading comprehension ability using the fan procedure to manipulate concurrent load (for a review, see Engle [26]).

Anderson [2,3] originally introduced the fan procedure to demonstrate that the speed and accuracy of retrieving information from LTM depends on the number of items to be retrieved. In the fan procedure, participants study a set of thematically unrelated propositions, with each proposition comprising two arguments (e.g., “The lawyer is in the boat”). The number of shared concepts for these propositions is varied to increase the memory set size (i.e., propositional fan size; e.g. “The lawyer is in the hall”). Once participants have learned a set of sentences, they are given a speeded verification task containing both the studied sentences and foils. The basic finding was that reaction times (RTs) and error rates increased as the fan size increased [2,3]. These results were thought to reflect processes related to the spread of activation among concepts in LTM, with greater propositional sets requiring more active concepts, which in turn led to longer retrieval times (ACT* model [3]).

In a WM framework, Cantor and Engle [13] originally assumed WM capacity to be identical with capacity for spreading activation in LTM. To test this hypothesis, they compared the effects of fan on individuals with low and high WM capacity in complex span tasks. They found the increase in RTs as a function of fan to be steeper in low- than high-capacity participants, supporting their hypothesis. A later study by Conway and Engle [17] contrasted retrieval time from primary memory and LTM as a function of set size. They found that the retrieval function from primary memory was steeper for low-WM-capacity individuals compared to high-capacity individuals, whereas there was no difference in the time needed to bring a set from LTM to primary memory. Importantly, they also found that the difference between low- and high-capacity individuals was present only in conditions that included competing information comparable to the shared arguments in the fan manipulation. When no overlap between different memory sets was present, WM capacity did not affect retrieval times. The authors concluded that the WM capacity measured by complex span tasks reflects executive attentional capacity needed to cope with keeping competing information in an accessible and discriminable state. The result suggests that the ability of complex span tasks to predict coping with distraction results from these tasks.
placing a heavy demand on capacity to activate sets of competing information. The present study addresses the question of how WM load of the fan type, in a task resembling complex span tasks, affects sentence comprehension as seen in an ERP response generally thought to reflect semantic comprehension.

1.2. ERPs and the semantic N400 response

Event-related brain potentials provide a noninvasive, on-line measure of neurocognitive processing. They are derived from scalp recorded electroencephalographic (EEG) activity and are obtained through a signal averaging process in which the EEG activity is time-locked to the onset of stimuli of interest and then averaged over experimental trials to extract the nonrandom electrical activity [63]. The main strength of ERPs lies in their exquisite temporal resolution, but it is also possible to derive spatial information about the generators (or sources) of ERP components. Typically, the spatial analyses of ERP (or magnetoencephalography, MEG) data require higher resolution electrode arrays (e.g., 64, 128, or 256) in order to obtain detailed topographical information and model the active sources [61].

A number of ERP studies on language have focused on the N400 component [49]. The prototypical N400 is elicited to semantically incongruent terminal words of visually presented contextually constrained sentences (e.g., “She takes her coffee with cream and dog”, when ‘sugar’ is expected). These semantic anomalies elicit a late negative-going peak at approximately 400 ms post-stimulus, which is predominately seen on the centro-parietal electrodes. A good deal of evidence suggests that the N400 can be elicited in a variety of different circumstances—all of which share a common emphasis on semantic evaluation. For instance, it has been observed using related and unrelated word pair stimuli [8,11], spoken sentences [15,39], and cross-modal paradigms with pictures and words [16,30,55].

With respect to the functional interpretation, there is evidence that the N400 reflects controlled post-lexical semantic integration processes. Holcomb [38] reported that degraded stimuli delayed the latency of the N400 regardless of whether the words were primed or unprimed. The results were interpreted in favor of post-lexical processing because there was no interaction between priming and stimulus degradation. That is, it was hypothesized that if the N400 reflected pre-lexical processing, then the latency delays resulting from stimulus degradation would be smaller in the primed than the unprimed condition. This prediction was not supported. Similarly, Brown and Hagoort [11] used masked and unmasked primes and found that the relatedness of the prime influenced the amplitude of the N400 only in the unmasked condition. If the N400 processes were pre-lexical, then preventing the stimulus from reaching conscious perception should not have affected the response. The finding that the primes modulated the N400 amplitude only when they were consciously perceived was taken to support the role of post-lexical semantic integration for the N400 response. However, the involvement of pre-lexical automatic processes in the generation of the N400 cannot be discounted yet, as evidence for pre-lexical processing has recently been reported [24].

Electromagnetic studies (i.e., ERPs or MEG) that localize N400 sources (in visual and auditory modalities) have begun to reach a general consensus with respect to some of the key generators. The middle and superior portions of the left temporal lobe, particularly in the posterior regions, are thought to provide major contributions to N400 scalp recorded activity [22,34,36,37,71]. In addition, homologous temporal areas in the right hemisphere, as well as polymodal regions in the left temporal-parietal-occipital region (BA 37) have also been implicated [22,34–36].

The anterior medial temporal lobe (AMTL) has been identified as another major contributor to N400 activity. McCarthy et al. [50] and Nobre and McCarthy [56] used intracranial electrodes in different brain areas and found N400 generators in the AMTL (including the inferior temporal neocortex, the anterior fusiform and parahippocampal gyri). Simos et al. [71] and Halgren et al. [35] replicated these results using MEG, localizing components of the N400 network in the vicinity of the AMTL. While the converging evidence is promising, it is still not possible to assume a single neural architecture for the N400. Haan et al. [34] used current density reconstruction algorithms (CURRY) and found a broad and scattered distribution of N400 sources within the cortex. In addition, they reported the existence of individual differences in topographies and cortical current source estimates. These results indicated that N400 source locations vary both within- and between-individuals.

1.3. ERP studies of WM

The effects of WM on ERPs have been examined in a number of studies [51,77]. Rössler et al. [64] used the fan procedure with concrete nouns to investigate tonic voltage changes in ERPs (which were thought to reflect the excitability of the underlying cortical tissue). They found that memory probes elicited DC-like negative slow wave potentials, particularly over the left anterior region, and increases in fan size produced potential changes over the frontal and parietal sites for negative probes (i.e., items not in the memory set). Some WM effects on the N400 have also been examined within the context of aging. Gunter et al. [33] used memory-demanding sentences to examine ERPs in younger and middle-aged participants. They utilized syntactic sentence structure (cf. King and Just [45]) to create sentences that varied in terms of working memory load and terminal word congruence. Increased WM load reduced the N400 congruency effect in the younger age group and eliminated the congruency.
effect in the middle-aged group. The congruency effect was also sensitive to individual differences in WM capacity during the processing of complex syntactic structures. The authors concluded that most ERP-effects, such as the N400, were dependent on WM capacity. There are two important points related to these N400 results: (1) within-subject WM effects appeared to be limited to a relative reduction in the magnitude of the N400 effect, *but not the timing*; and (2) the results are best represented as a reduction in the ‘congruency effect’ because WM load reduced the relative difference between incongruent and congruent conditions (and not necessarily the peak amplitude of the N400 in the incongruent condition).

1.4. Objectives and hypotheses

The present study attempted to characterize the specific influences of WM load on the N400 using a fan procedure modified for ERP recordings [2,3]. The fan procedure was selected because it is well characterized in behavioral research, allowing for WM effects on the N400 to be interpreted within an established theoretical framework. The paradigm was adapted from a previous visual-to-auditory sentence-matching task used to investigate speech perception [22]. The modified paradigm used a visual-to-visual sentence-matching task. The sentence pairs contained ‘prime’ and ‘target’ sentences that were characterized by an exemplar-superordinate semantic hierarchical relationship. For example, for the prime sentence “The man is climbing into the cockpit”, the target sentence was “The man is in the airplane”. The prime sentences comprised a subject (man, woman, boy, or girl), a verb, and a preposition phrase describing a location. They were constructed to establish an expectancy, which was evaluated subsequently within the context in following target sentences. In the target sentences, the last words either matched or mismatched the expected superordinate location. Mismatching or incongruent terminal words had, in a previous speech perception study, been found to elicit the N400 [22]. In the present study, WM load (and fan) was manipulated by presenting either one priming sentence (WML1) or two priming sentences (WML2) and a target sentence. In WML2, two possible endings consistent with the two priming sentences have to be kept in active working memory to perform the task of deciding whether the terminal word matches or mismatches expectation. As the two priming sentences shared their subject noun phrase, the ending locations created a fan of two facts about the shared subject.

It was predicted that the behavioral fan effect would be present in the RT data using the modified fan procedure, that is, decision times would be longer in the WML2 condition (Hypothesis 1). Increased WM load was expected to reduce the congruency effect, in line with previous research and suggesting an interaction between WM and sentence processing. It was also predicted to delay the N400 peak latency, indicating a WM search before semantic resolution of the target sentence in the WML2 situation (Hypothesis 2). High-resolution ERPs were used to localize the N400 sources. Therefore, we also explored WM changes in activation within the N400 source models (and the source waveforms).

2. Methods

2.1. Participants

Sixteen university students (10 females and 6 males) volunteered to participate in a study on language and memory (course credit was given where applicable). One additional participant was tested, but the data were not included in the study due to self-reported reading difficulties. All 16 participants were fluent in English, their mean age was 22.1 years (S.D.=2.5), and their mean level of education was 15.9 years (S.D.=1.3). The Edinburgh Handedness Inventory was used to assess handedness and all participants were dextral (LQ range=39.1–100 [58]). All had normal or corrected-to-normal vision and were screened using a self-report measure for audiological, psychiatric, and/or neurological history. The study had ethical committee approval and informed consent was obtained before the experiment proceeded.

Participants’ simple span capacities were assessed using the Digit Span (Forward, Backward, and Total) from the Wechsler Memory Scale-Revised (WMS-R [78]). Their mean scores for Total, Forward, and Backward were 17.2/24 (S.D.=3.1, range=12–24), 10.1/12 (S.D.=1.5, range=7–12), 7.1/12 (S.D.=2.2, range=4–12), respectively. Using a split-half procedure, High and Low groups were derived from the Total scores. The mean score for the High group was 19.5 (S.D.=2.3) and the mean score for the Low group was 14.9 (S.D.=1.8).

2.2. Experimental materials

Two hundred sentence pairs were constructed in the form “The SUBJECT is VERB-ing in the LOCATION” “The SUBJECT is in the...?”. The sentence pairs were then given as a normative survey to 108 individuals (72 females and 36 males; all fluent in English). The sample mean age was 27.5 years (S.D.=10.2) and the mean level of education was 17.1 years (S.D.=3.3). A subset of 180 was selected from the 200 sentence pairs, only items with error rates greater than 33% or repeat words and sentences were excluded. The semantic congruence of the terminal words in the target sentences was manipulated to match or mismatch expectation. Congruent terminal words were the highest probability words for each sentence pair [9]. Incongruent terminal words were semantically unrelated words with initial phonemes that did not match any of the
normative responses. The incongruent words were selected from alternate words in the normative data set to control for word frequency effects. All conditions were equal with respect to probabilities and word lengths.

The sentence pairs were divided into two levels of WM load (WML1: 60 pairs and WML2: 120 pairs). WML1 contained a single prime and target sentence (e.g., Prime: The woman is riding on the underground train. Target: The woman is in the subway/church). WM load was increased in WML2 by using two unrelated sentences in the prime (e.g., Prime: The boy is sitting on the witness stand. The boy is standing at the grave. Target: The boy is in the courtroom/laboratory/cemetery). WML1 and WML2 were both comprised of 60 trials (30 congruent and 30 incongruent trials). The stimuli were presented in a fixed pseudo-random and counterbalanced order (WML1: WML2 and WML2: WML1). Given the high trial numbers that were required for signal averaging (which increased exponentially with memory set size), the study was limited to set sizes of one and two propositions.

The prime and target sentences were presented visually on a 14-in. computer screen positioned 1 m from the participant’s nasion. The text stimuli were presented in yellow letters on a black background (36-font size). For WML1, a single prime sentence was presented in the center of the computer screen. For WML2, two prime sentences were presented simultaneously in the top and bottom half of the screen. WML1 prime sentences had a mean width of 19.76 cm (S.D.=2.13) and mean height of 3.00 cm (S.D.=0.62). WML2 prime sentences had a mean width of 20.83 cm (S.D.=1.87) and a mean height of 11.02 cm (S.D.=0.55). The prime sentence stimuli subtended a maximum visual angle of approximately 14.3°. The prime sentences were presented at once and target sentences were presented in rapid-serial-visual-presentation (RSVP) format, one word at a time. Target sentence words subtended a maximum visual angle of approximately 5.7°. ERPs were recorded from the onset of the terminal words in the target sentences.

2.3. Procedure

Digit Span (Forward and Backward) from the Wechsler Memory Scale-Revised (WMS-R [78]) was administered to all participants prior to the experiment. The ERP experiment began with a practice phase, comprising the task instructions and sample trials. The prime stage (containing either one or two sentences) was presented for 10 s. The target stage (containing one sentence) was presented with a 0.5-s duration per word (with negligible time between words). A 6-s interval defined the time between the terminal word in the target sentences and the onset of the next trial (with a 3-s behavioral response window). Participants pressed a button to indicate whether the terminal words matched or mismatched expectation. They were told to respond as quickly and as accurately as possible (speed/accuracy trade-off) and their response hand (left or right) was counterbalanced between subjects. During the recording session, participants were told to remain as still as possible and to try not to blink during the terminal word presentation.

EEG activity was recorded using a 128-channel NeuroScan Synamps™ system. The electrodes were embedded in a NeuroScan Quik-Cap™, positioned on the basis of an extension [1] of the International 10/20 System [42]. Vertical and horizontal eye movements as well as blinks were recorded using an electrooculogram (EOG), with electrodes above, below, and on the outer canthi of both eyes. The impedances were maintained at or below 10 kΩ and AFz served as a ground electrode. All electrodes were Ag/AgCl and an electrode located on the nose served as the reference. The continuous EEG recordings (band pass: 0.05–30 Hz, digitally sampled at 500 Hz) were epoched off-line (−100–1000 ms) and digitally low pass filtered at 20 Hz. Any trial with EOG artifacts greater than ±75 μV (−100–750 ms) were excluded from the analysis. The mean percentage of trials accepted in the analysis was 96.5 (S.D.=4.3) for WML1 and 91.1 (S.D.=8.8) for WML2. The remaining trials were averaged by experimental condition and only trials in which the individual provided a correct response were included in the average The mean percentage of correct responses was 94 (S.D.=9.5). The mean number of trials per condition was 27.9 (S.D.=1.5) in WML1-Congruent, 28.9 (S.D.=1.7) in WML1-Incongruent, 24.0 (S.D.=6.7) in WML2-Congruent, and 24.3 (S.D.=6.8) in WML2-Incongruent. Subtracting the Congruent condition from the Incongruent condition derived difference waveforms. Difference waveforms were used to isolate N400 activity. Individual waveforms were also averaged together to create grand average waveforms.

Statistical analyses were conducted on both the behavioral and electrophysiological data. The analyses were conducted using a repeated measures analysis of variance (ANOVA) with conservative degrees of freedom [32]. Significant main effects and interactions were submitted to additional post hoc analyses using the Tukey Honestly Significant Difference (HSD) test. An alpha level of $p<0.05$ was required for statistical significance. Accuracy and reaction time (RT) data were analyzed separately, with the factors being Order (two levels), WM load (two levels), and Congruency (two levels). Before deriving the individual mean RTs, trials with values greater than 2 s, outliers ($±2$ S.D.), and errors were excluded from the analysis. The main ERP factors were Order (two levels), WM load (two levels), Congruence (two levels), and Region (nine levels). The number of statistical comparisons across electrode sites was reduced to nine scalp regions using a linear derivation to recompute the waveforms. The electrodes were separated laterally into left (L), midline (M), and right (R) sectors. The sectors were also subdivided into
frontal (F), central (C), and posterior (P) sectors to obtain the nine regions (LF, MF, RF, LC, MC, RC, LP, MP, and RP). Ten electrode sites were used to compute the waveform for each region. For two individuals (S02 and S13), the MP and MF sections (respectively) contained data from only nine electrodes because noisy electrodes were excluded.

Source analysis was conducted using Brain Electromagnetic Source Analysis (BESA, MEGIS version 3.0 [66]) and is necessarily an approximation. Difference waveforms were examined to isolate N400 activity and limit the number of dipoles in the solution. WML1 and WML2 were modeled separately because it was predicted that brain activation changed as WM load increased. Due to the intensive nature of the analysis, solutions were obtained using grand average waveforms over subjects [61,67,68]. Grand average models provide a global estimate of the active regions across individual data sets. Individual variance in source localization was evaluated to determine whether the fit was reliable. An overview of the modeling procedure is provided below.

2.3.1. Pre-processing

Channel labels as well as positional values (theta and phi) were imported into BESA. Additional electrode locations were determined using the 10/10 system electrode landmarks. The data were pre-processed further using a more conservative artifact rejection procedure to minimize ocular artifacts (−100–1000 ms; WML1 accepted trials M=91.0%, S.D.=14.9; WML2 accepted trials M=85.1%, S.D.=14.8). As a standard procedure in BESA, the averaged data (0–1000 ms) were also refiltered with the low pass setting at 10 Hz and residual EOG artifact activity was accounted for by using an ocular regional source. Artifact channels and electrodes contaminated with noise (e.g., electrodes over the ears) were excluded from the analysis.

2.3.2. Modeling

The dipole modeling procedures were the same as those used in a prior work [22]. Briefly, a principal component analysis (PCA) was used to estimate the number of sources, assuming that the decomposed source activities were not independent. PCA revealed that five or six bilateral sources accounted for 98–99% of the variance (10–12 dipoles total). A regional source strategy was selected to localize the dipoles and identify the source waveforms [67]. In the initial model, regional sources (comprising three orthogonally oriented dipoles in the same location) were used to identify active areas of current flow. To begin, bilateral sources (symmetric across the hemispheres, df/source=3) were placed in established N400 regions within the temporal lobes [34–36,50]. Additional sources were added if required and the locations were fitted separately and together until the model achieved stability. Once active regions were identified, releasing the regional constraints (location and orientation) and fitting the individual dipoles (separately and together) formed an advanced model. Bilateral location constraints were used throughout the fitting procedure because releasing them did not improve the model. Improvements to the model, either through additional dipoles or changes in location/orientation parameters, were evaluated primarily on the basis of reductions in the residual variance (RV: a minimum 0.5% reduction was required). Orientation and location information provided the estimates for neuroanatomical regions. A minimal energy criterion (20% of cost function) was used in the fitting procedures in order to minimize source interactions (i.e., linear dependency). The validity of the model was evaluated primarily on the basis of RV (80% of cost function), as well as projected spline maps and prior physiological/anatomical evidence. Response-related dipoles were included within the model for additional physiological validation. The RV for the 303–505 ms interval-of-interest was then recorded to evaluate the model’s goodness of fit (the inverse of RV).

3. Results

3.1. Behavioral analyses

The objective of the behavioral analyses was to determine whether the fan effect was replicated (Hypo-
The accuracy data were analyzed using a repeated measures ANOVA with Order (two levels), Congruency (two levels) and WM load (two levels) as factors. There was a significant main effect of Congruency, $F(1,14)=5.70$, $p<0.05$, indicating that the participants’ performance was better for Incongruent words ($M=95.2\%$ correct, SE=1.6) than for Congruent words ($M=92.6\%$ correct, SE=2.1). There were no other significant main effects or interactions. The RT data were analyzed using repeated measures ANOVA with the same factors. There was a significant main effect of Order, $F(1,14)=5.21$, $p<0.05$, with shorter RTs in the WML1:WML2 Order ($M=740$ ms, SE=71) and longer RTs in the WML2:WML1 Order ($M=970$ ms, SE=71). The main effect of Congruency was also significant ($F(1,14)=30.28$, $p<0.0001$) with shorter RTs in the Incongruent condition ($M=840$ ms, SE=50) and longer RTs in the Congruent condition ($M=870$ ms, SE=51). Importantly, there was a significant main effect of WM load, $F(1,14)=10.55$, $p=0.01$, indicating that the RTs in WML1 ($M=781$ ms, SE=49) were shorter than those in WML2 ($M=929$ ms, SE=61). The interaction of interest between the Congruency and WM load factors was also significant, $F(1,14)=5.91$, $p<0.05$. Subsequent post hoc analyses revealed that WML1 RTs were shorter than WML2 RTs, with the Incongruent condition showing the maximum difference between WML1 and WML2 (~168 ms delay). There were significant Congruency differences (Incongruent>Congruent; ~50 ms) in WML1, but no differences existed in WML2 (~10 ms difference). No other interactions were significant.

In order to evaluate the fan effect between the High and Low groups, the RT data were reanalyzed using the between-subjects factor WM group (High and Low). The same within-subjects main effects and interaction were again significant following the reanalysis (i.e., Congruency, WM load, and Congruency x WM load; all $F>5.0$ and $p<0.05$; replicating the prior result). In addition, there was a significant WM group x WM load interaction, $F(1,14)=5.46$, $p<0.05$. Post hoc analyses revealed that the fan effect for the Low group was larger than that of the High group. Fig. 1 shows the behavioral fan effect for the RT data.

---

![Grand average ERPs](image.png)

Fig. 2. Grand average ERPs ($N=16$) to terminal words in the Congruent and Incongruent conditions for WML1 and WML2. Nine regions across the scalp are presented (10 electrodes per region). EOG sites were artifact free (data not shown). A N400 component was elicited in both Incongruent WML1 and WML2 conditions. Time (ms) is on the x-axis and voltage ($\mu$V) is on the y-axis (negative is up).
3.2. Electrophysiological analyses

The main objective of the ERP analyses was to evaluate the effects of Congruency and WM load on the N400. Fig. 2 presents grand average waveforms for the Congruent and Incongruent conditions at WML1 and WML2. The initial analysis examined the standard waveforms with Order (WML1:WML2 and WML2:WML1), Congruency (Congruent and Incongruent), WM load (WML1 and WML2), and Region (LF, MF, RF, LC, MC, RC, LP, MP, and RP) as factors. Follow-up analyses examined more closely the effects of WM on the amplitude and latency characteristics of the N400 using difference waveforms.

3.2.1. N400 standard waveform analysis

The initial ANOVA was conducted on the 350–450 ms time interval (average amplitude method). There was a significant main effect of Congruency ($F(1,14)=107.73$, $p<0.0001$) with larger negative-going amplitudes in the Incongruent condition ($M=-5.99$, $SE=1.35$) compared to those in the Congruent condition ($M=5.99$, $SE=1.35$). There was also a significant main effect of WM load ($F(1,14)=24.60$, $p<0.0001$) resulting from more positive amplitudes in WML1 ($M=3.86$, $SE=1.15$) as compared to WML2 ($M=1.54$, $SE=1.26$). There were no other significant main effects. The Congruency × WM load interaction was significant, $F(1,14)=11.49$, $p<0.005$. Subsequent post hoc analyses revealed that the difference between the Incongruent and Congruent conditions was reduced as WM load increased. Importantly, while the amplitudes in the Incongruent condition did not differ between WML1 and WML2, the positive amplitudes in the Congruent condition at WML1 were significantly reduced at WML2. The interaction revealed that increased WM load reduced the positive-going response in the Congruent condition. No other interaction was significant.

3.2.2. N400 difference waveform analysis

To examine latency effects of WM load on the N400, difference waveforms were derived by subtracting the Congruent from Incongruent (WML1 and WML2). The difference waveforms depict the changes in N400 peak latency as a function of WM load. There was a reduction in peak amplitude (due to changes in the Congruent condition) and a delay in peak latency. The nine regions show the Midline Central (MC) distribution of the N400. All other details as for Fig. 2.

---

1. Additional analyses were conducted on the N200 time interval. For the sake of brevity, these results are not reported as they demonstrated only an Order effect on the N200 amplitude (i.e., Order did not affect the N400).
with WM Load (two levels) and Region (nine levels) as factors).

Fig. 3 depicts the N400 in difference waveforms for WML1 and WML2 at all nine regions. The amplitude analysis confirmed the results of the prior ANOVA, with a significant main effect of WM load \((F(1,15)=9.2, p<0.01)\). There was also a significant main effect of Region \((F(8,120)=2.74, p<0.05, \eta^2=0.469)\) and subsequent post hoc analyses revealed that the N400 peak was largest in the MC region. The WM load × Region interaction was not significant \((F<1.0)\).

With respect to peak latency, there was a significant main effect of WM load, \(F(1,15)=11.21, p<0.005\). The peak latencies for WML1 \((M=397.6 \text{ ms}, \text{SE}=10.35)\) were shorter than peak latencies for WML2 \((M=448.4 \text{ ms}, \text{SE}=11.24)\), demonstrating a significant N400 latency delay resulting from increased WM load \((\text{mean delay: }50.8 \text{ ms})\). Neither the Region main effect nor the WM Load × Region interaction was significant \((F<1.0)\).

### 3.2.3. Individual N400 peak latency delay

Because individual differences in WM capacity have proven to be important in behavioral work [26], the relationship between individual WM differences and N400 peak latency delays was explored further. Grand average waveforms were also examined for high and low subject groups. Subjects with lower WM capacity showed more pronounced latency delays. In order to explore whether a relationship existed, a correlation analysis was done to determine whether the latency delays were significantly related to performance on the Digit Span. For each participant, the N400 peak latencies were obtained for WML1 and WML2 using the MC regions from both the difference and standard waveforms. Subtracting the latency values at WML2 from those at WML1 derived the peak latency delay.

Examination of the latency delays using difference waveforms revealed no significant correlations with individual WM capacity. However, a correlation analysis of latency delays in the standard waveforms revealed a significant inverse relationship (two-tailed) with performance on Digit Span. N400 peak latency delays correlated negatively with the Total scores \((r=-0.52, p<0.05)\) and, more importantly, they also correlated negatively with the Backward scores \((r=-0.55, p<0.05)\). However, the latency delays did not correlate with the Forward scores \((r=-0.27, p=0.315)\).

### 3.3. Source analysis

The objective of the source analysis was to evaluate WM effects on the N400 sources. The specific goals were: (1) to localize the N400 sources separately in WML1 and WML2 (confirming the modeling process); and (2) to examine whether WM load differences existed between the two models. Fig. 4 depicts the source models for WML1 and WML2. There was a close overlap between the N400
sources localized in WML1 and WML2, with active bilateral areas in the lateral occipital gyri (N400S1: BA 18), the occipital-temporal regions (N400S2: BA 19), and the temporal poles (N400S3: BA 38). An additional source was localized to the left superior temporal gyrus or STG (N400S4: BA 22). With respect to WM load, the major difference between the two models was novel source activation in the left inferior parietal lobe (N400S5: BA 40) for WML2 only. The activation associated with N400S5 was consistent with the notion that this region supports working memory [12,53,60].

Fig. 5 shows the source waveforms for WML1 and WML2 models. Examination of source waveform data revealed that the peak activation occurred in the N400 time range (404 ms). Importantly, there were also identifiable differences in the activation between the two levels of WM load. A number of sources (N400S1,S2,S4) showed more pronounced peak activation in the N400 time range at WML1. In contrast, the anterior temporal and left parietal sources (N400S3,S5) were characterized by enhanced activation in WML2. A probe dipole in the left inferior parietal lobe region revealed no signs of activation in WML1. Overall, the changes in source activation between the two levels of WM load reflected increased mnemonic processing in the anterior temporal and left parietal sources.

Both the WML1 and WML2 models accounted for more than 94% of the variance in the N400 time interval-of-interest (303–505 ms). Specifically, the RV for WML1 was 5.03% and the RV for WML2 was 5.86%. The projected topographical maps for the N400 matched closely the data maps and physiological validation of the models was provided by the localization of response dipoles (half of the participants responding with their right and half with their left hand) in the primary motor and sensory cortices. Similar to previous work [22], these sources accounted for activation associated with the participants’ button press responses following terminal word onset. To evaluate whether the source models were representative across individuals, both WML1 and WML2 initial models were tested using individual data. The source locations fit the individual data sets reliably. The mean RV for WML1 was 12.2 (SE=1.5, range 4.96–26.3) and the mean RV for WML2 was 25 nAm.

![Visual-perceptual and semantic processing](image)

![Memory processing](image)

Fig. 5. A direct comparison of the source waveforms between WML1 and WML2. The source waveforms are also divided on the basis of the functional task requirements. Activation associated with evaluating visual-perceptual and semantic processing is observed in the bilateral occipital, bilateral occipital-temporal, and left STG areas (top). In contrast, sources in bilateral temporal poles and left inferior parietal region reflected activation associated with the task-related memory demands (semantic retrieval and working memory). The probe source represents activation for the left parietal region at WML1. All data are presented on a uniform scale with Time (ms) on the x-axis and Voltage (nA) on the y-axis.
was 12.4 (SE=1.1, range 7.69–24.2). The results showed that on average both models accounted for more than 87% of the variance for the individual subjects.

4. Discussion

4.1. The behavioral fan effect and the N400 response

Analysis of the behavioral RT data revealed the presence of a fan effect (Hypothesis 1). Specifically, the increase in WM load resulted in a corresponding increase in RTs and this effect was most pronounced in participants who had lower WM capacity. It should be noted that the modified fan procedure differed from typical fan procedures in two main respects. First, participants were not required to study a proposition set prior to the experiment because the paradigm utilized semantic hierarchical associations derived from general knowledge. This difference may have accounted for shorter RTs and higher accuracy to Incongruent versus Congruent stimuli at WML1. Given that a number of possible congruent options exist (e.g., ‘classroom’ activates ‘school’, ‘college’, ‘institute’ et cetera) [22], it was likely easier to reject an incongruent item than to accept a congruent item when WM load was low. With two sentences, WM load increases and there is competition between the two sentences (as two possible locations must now be maintained in WM). As a result, it is less likely that multiple candidates for a single proposition are activated (e.g., ‘classroom’ activates only ‘school’), therefore reducing the difference in Incongruent/Congruent performance. Second, the task was limited by methodological imaging constraints to memory set sizes of either one or two propositions. In contrast, traditional fan procedures have used at least three set sizes (e.g., 1, 3, and 4). While the data do not allow us to separate a true fan effect from a simple load effect and are limited to set sizes of one and two propositions; the preliminary results are promising. Indeed, the RT delay (~168 ms) falls well within the range found in WM tasks with fan (~150–200 ms; e.g., Cantor and Engle [13]). Thus, it is reasonable to predict similar RT trends in this task when larger propositional set sizes are used in future research.

Increased WM load reduced the congruency effect and delayed the latency (Figs. 2 and 3). The reduced congruency effect was consistent with the Gunter et al. [33] findings. In both studies, there was a prominent reduction in the relative difference between the Incongruent and Congruent conditions, which resulted from a smaller positive-going component in the Congruent condition as WM load increased. While there was no significant change in the N400 peak amplitude in the Incongruent condition (Fig. 2), it is not clear if this activation interacted with on-going slow waves changes, which have been shown previously to be sensitive to WM load [51,64]. In contrast, there were amplitude changes in the Congruent condition, indicating that WM demands altered recognition processes related to the evaluation of competing semantically acceptable words. One interpretation is that the reduced congruency effect was due to an attenuation (or delay) of the positive-going activation that reflects the processing of acceptable sentence endings (i.e., a P300 component). Another related interpretation would be that increased WM load produced, in fact, a small N400 as a result of having two competing congruent endings rather than one highly expected ending. Further work is needed to delineate the different cortical potentials that contribute to WM changes in N400 congruency effect [51].

With respect to timing, a significant N400 latency delay (~50 ms) was present in the Incongruent condition (Hypothesis 2, Fig. 3). This finding provides physiological evidence for early delays in the temporal nature of semantic processing due to increased WM load. It is likely that these early delays ultimately result in, or at least contribute to, the behavioral fan effect (as measured by RTs). This result also represents the first evidence for memory-based influences on the N400 latency. Interestingly, the 50-ms delay fits remarkably well within the original Sternberg [75] estimates of the scanning slopes for active memory (36 ms for digits, 45 ms for nonsense forms, and 56 ms for faces) in a task requiring simple matching of probes with memory set members. Given that the N400 in this study reflected processes linked to controlled post-lexical semantic integration [47], it is possible that the latency delay results from the scanning of the activated set of congruent endings. In support of this interpretation, there were preliminary indications that individual latency delays may be related to individual WM capacities. Search rate in active memory has been suggested to contribute to individual differences in simple span tasks, like digit span forwards [41], as well as more complex span tasks [12], like backward span. According to an interpretation by Conway and Engle [17], the relation between comprehension and WM capacity depends on resources available to inhibit distracting information. In the present version of the fan procedure, the distraction consisted of two competing propositions with different ending words being available to describe the same subject. A fast N400 would result from a quick rejection of both congruent endings when faced with an incongruent ending. Individual differences in the speed of coping with the two congruent competitors to an incongruent ending could reflect individual differences in WM resources. Provided that further research confirms and elucidates the relationship between N400 latency and WM load, the results may shed light on theories about individual differences in WM capacity and reading comprehension [26]. Extensions of this work should include dual task measures that are more sensitive to individual differences in the kind of WM capacity relevant for coping with interference (e.g., reading span [19]). In particular, it would be interesting to investigate whether N400 latency-delay in different load conditions can be used as a physiological factor that predicts an individual’s WM capacity in a complex span task.
Source analysis, which is necessarily a spatial approximation, was used to explore changes in WM activation in the dipole models (with less emphasis on spatial resolution). It was possible to localize N400 generators to brain areas associated with processing semantic congruencies and memory, and to show that increased WM load led to novel source activation within this network. The localized sources for N400 activity corresponded to previous findings in the literature and explained a high percentage of the variance (>94% for both models). Given the large response (5–10 μV), the reliable goodness of fit (grand average and individual data), and the separate confirmation (WML1 and WML2), the results provide a good estimate of the network that supports the N400 elicited in this task. Moreover, the spatial estimates are highly consistent with previous N400 source localization work. In terms of semantic comprehension, the left perisylvian region has consistently been identified as one of the primary contributors to the N400, for both visual and auditory studies, using both ERPs and MEG [22,35–37,71]. This result is also consistent with PET and fMRI studies on language in general [12,25,62]. In addition, it appears that in the visual modality, activity related to the N400 may also be extended to include occipital-temporal areas, possibly involving the ventral visual stream [46,52]. These areas have also been suggested to play a role in automatized reading [65,70]. Perhaps the most interesting result with respect to the source analysis was the WM related changes in source activation that occurred in the left inferior parietal lobe and the bilateral anterior medial temporal lobes (AMTL). These results are discussed in more detail below.

4.2. Partitioning WM: insights from neuroimaging data

In the current investigation, the theoretical framework proposed by Baddeley [4,5] has been adopted for the more detailed interpretation of the results because it is the prominent model for which most neuroimaging results have been interpreted [7,72,74,76]. Since the conception of their original model, Baddeley and Hitch [6] have regarded WM as a system with specialized components that include a supervisory system (the central executive), as well as temporary memory systems for storing phonological (the phonological loop), and visuospatial (the visuospatial sketch pad) information [4,5]. The present findings suggested that WM influences in a sentence-processing task involving response competition were attributable to neurocognitive systems that support the phonological loop and, to some degree, the central executive. Examination of the source waveform data revealed sources with a clear mnemonic role (Fig. 5). The key active areas included the bilateral temporal poles (N400S3) and the left inferior parietal lobe (N400S5) (Fig. 4).

The anterior temporal sources may have been related to semantic memory retrieval and were comparable to the results of intracranial studies, showing N400-like activation in the AMTL [50,56]. Nobre and McCarthy [56] proposed a mnemonic role for the AMTL N400, attributing it to either activity in a network of semantic representations associated with the word or post-lexical integration of the word into ongoing context. PET and fMRI reports of semantic retrieval activation in this region have been scarce, perhaps attributable to technical difficulties associated with imaging function in this region [12]. Nonetheless, there is some recent evidence for left anterior medial temporal (hippocampal and parahippocampal) semantic activation in an event-related fMRI studies [44,54].

In contrast, a number of PET and fMRI studies have identified the left parietal region as a primary contributor to verbal working memory [60]. ERP evidence for verbal WM activation in this region has also now been revealed during speech perception [22]. Further, it has been shown that prolonged RTs in a verbal WM task were linked specifically to corresponding increases in the power of functionally active parietal regions [40]. This pattern matches the current RT trend and activation profile. The phonological loop has been subdivided into a passive phonological store and an active rehearsal system [4,5]. It has been suggested that the parietal lobe plays a central role in the phonological store. The present findings therefore suggest that phonological WM resources are recruited to help the central executive cope with the complex demands of the present sentence-processing task.

Becker et al. [7] have raised important concerns about the need for a finer grained analysis of the foci of the phonological store within the parietal lobe. In a review, the authors suggested that there might be different foci, none of which are fully compatible with the theory of the phonological loop. Becker et al. [7] proposed two criteria for defining the location of the putative phonological store: (1) the locus of activation must be reliable and functionally parsimonious (i.e., appropriately located and activated by verbal WM tasks); and (2) the locus must be sensitive to a variety of tasks that may activate the store (e.g., convergence from active and passive verbal tasks), but still specific to verbal WM (i.e., not active in unrelated tasks).

Increased WM load led to novel N400 source activation in the left inferior parietal lobe. These results agree with Becker et al.’s reliable identification of this locus (first criterion), but the authors stated that because the area is not typically identified for speech perception and has also been active in visual attention tasks—it does not satisfy the second criterion (i.e., sensitivity and specificity). Comparable left inferior parietal ERP findings during speech perception (BAs 39/40 [22]) may support the notion that WM activation in speech perception and reading tasks converges on a unitary store. However, it is still too early to conclude that a single locus is active across modalities because of the spatial resolution limitations of BESA (and source analysis in general). Therefore, the locus of the putative phonological store remains to be determined.
It could be thought that one perplexing aspect of the present results is the lack of localized sources in the dorsolateral pre-frontal cortex, often implicated in fMRI studies [43]. The source analysis focused specifically on equivalent current dipoles that showed peak activation in the N400 time range (i.e., those with phasic changes of electromagnetic potentials). During the source fitting procedure, local minima (i.e., regions that accounted for waveform variance) were detected in the lateral prefrontal and parietal regions, bilaterally. Dipoles were not fitted in these regions because BESA cannot model low frequency signals accurately [61]. Nonetheless, this DC-like activation is reminiscent of prior work that used the fan procedure to study slow ERPs. Rösler et al. [64] found that memory probes elicited tonic voltage changes associated with the relative excitability of underlying cortical tissue in the left anterior region as well as frontal and parietal electrode sites in general. Indeed, we have recently investigated fMRI correlates of verbal WM using this paradigm and found activation in both parietal and frontal regions [23]. These findings concur with a number of PET and fMRI studies on WM that have used N-back tasks [57,72], which are similar to the modified fan procedure with respect to WM load manipulations. N-back tasks involve both short-term maintenance and the central executive for attentional control. In general, regions that are commonly implicated during these tasks include (but are not limited to) the prefrontal, parietal, and cingulate areas [10,69,72,73]. Taken together, involvement of lateral prefrontal [59] and parietal [40] regions in central executive functions and controlled attention may also have a central influence on semantic neural systems. Further work is needed, however, to elucidate the role of frontal areas in semantic processing.

4.3. Summary

The present study examined the effects of WM load on a complex task involving sentence processing using the semantic N400 response as the primary dependent measure. WM load was seen to reduce the congruency effect, suggested to result from response competition in WM, and delayed the N400 latency, suggested to be a consequence of fast checking procedures in the activated set of possible sentence endings. Localization of load sensitive sources for the N400 suggested activation in semantic memory networks, as well as recruitment of the phonological store to cope with an increased WM load in the task combining storage and processing of verbal material.

Acknowledgements

This work was funded by the Canadian Foundation for Innovation, the Natural Sciences and Engineering Research Council of Canada, the Scottish Rite Charitable Foundation of Canada, the Canadian Language and Literacy Research Network, University of Helsinki research funds, and the Academy of Finland (Grant 39253). The authors thank J. Bennett and E. Mazerolle for assistance with preparation of this manuscript.

References


