Published in final edited form as:

Psychophysiology. 2012 June ; 49(6): 770–785. doi:10.1111/j.1469-8986.2012.01366.x.

Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension

Edward W. Wlotko and

Beckman Institute, University of Illinois

Kara D. Federmeier

Department of Psychology, Neuroscience Program, and Beckman Institute, University of Illinois

Abstract

Contextual information influences multiple aspects of language comprehension extended over time. To determine how age-related changes impact normal comprehension, effects of contextual strength were examined with event-related potentials. Increased contextual constraint facilitated semantic processing (reduced N400s). Effects were smaller and delayed for older adults, and sensitivity to contextual information was diminished for weak contexts. Both groups elicited a later left-lateralized frontal negativity associated with reinterpretation of context when multiple interpretations of a sentence were likely. Older adults evidenced the frontal negativity over a wider range of constraint. The change in the use of contextual information across age is attributed to decreased reliance on predictive processing for older adults. Thus, age-related changes in comprehension lead to differential engagement of processing resources over time for older adults.

Comprehending individual sentences requires decoding word-level information from the spoken or written input, using that information to fit the word into its prior context in a combinatorial manner at semantic and structural levels, and bringing to bear inferential, situational, and background knowledge to further refine the meaning presumably intended by the writer or speaker. Thus, multiple mechanisms at several processing levels are engaged at different times in the comprehension stream to effect normal language comprehension, and these processes are significantly influenced by the amount and type of available information from the larger language context and the immediate environment (e.g., Simpson, 1991; Schwanenflugel & LaCount, 1988; Kamide et al., 2003; Brown-Schmidt et al., 2008; Hagoort & Van Berkum, 2007).

The examination of context effects in comprehension has been greatly aided by the use of event-related brain potential (ERP) methods, as these provide direct, specific, and multidimensional measures of neural activity over time, concurrent with cognitive processes of interest. In particular, many ERP studies of context effects have focused on the N400 component, which is a reliable measure of semantic processing. The N400 is elicited by any potentially meaningful stimulus and is modulated by factors that affect semantic processing, including the relationship of a particular stimulus to its preceding context (see Kutas &

Federmeier, 2011). N400 amplitudes are reduced when demands on semantic processing are decreased. For example, over the course of a congruent sentence, as contextual information increasingly constrains the upcoming words, semantic processing of those words is facilitated, with corresponding reductions in the amplitude of the N400 (Van Petten & Kutas, 1990). Across different sentences, N400 amplitudes vary with the amount of contextual information (sentence “constraint”) and the fit of a particular word into that context (Kutas & Hillyard, 1984). Sentence constraint and the contextual fit of a word into a sentence have both generally been operationalized using measures of cloze probability -- that is, the empirically determined proportion of individuals who agree on a completion for a particular context (Taylor, 1953). Strongly constraining sentences are those that lead most people to complete the context with one or only a few different words (which then have high cloze probability). Sentences that elicit only low cloze probability endings provide weaker constraints, allowing them to be completed in many ways; thus, words in weakly constraining sentences are less supported by those sentence contexts. Kutas & Hillyard (1984) demonstrated that N400 amplitude is powerfully predicted by cloze probability, in an inversely linear manner: at the group level, the correlation is around .9.

The N400, in combination with other ERP measures, has accordingly been used to great effect to determine which aspects of context are brought to bear on word processing, and when and how that contextual information is utilized. Traditional models of language comprehension have often posited that word recognition processes (linking the physical signal from spoken words or written text to information about word forms) occur via stereotyped procedures that are invariant across different types of context (see, e.g., Forster, 1981; Swinney, 1979). In this type of model, then, context effects occur only later in processing, when the accessed meaning of a word is “integrated” with the current message-level representation. The large body of ERP work – relying on measures concurrent with the cognitive processing under investigation rather than a downstream behavioral response – suggests instead that context use is incremental and often immediate (Van Petten & Kutas, 1991; see also DeLong et al., 2005; Van Petten et al., 1999) and that context must be broadly defined, as N400s are affected by information not only at the sentence level but also by discourse information (Van Berkum et al., 1999) as well as by extra-linguistic context (review in Van Berkum, 2008).

Furthermore, work using the N400 has been important in providing evidence that contextual information can be exploited in an anticipatory manner, preactivating information about likely upcoming words (review in Federmeier, 2007). The question of when predictive processing is brought to bear in comprehension can be difficult to answer because it is often the case that similar patterns would be expected whether contextual use is anticipatory or only exerts influence in late stages of word processing. That is, a strong context can facilitate processing because upcoming words are predictable, or because the context provides supportive information for the word to be integrated with after being recognized. ERP studies have been particularly instrumental in dissociating these accounts by looking for effects of prediction preceding or without presenting the critical predictable word. Many studies across several languages have shown that context can be used to preactivate information about likely upcoming words at semantic, syntactic, phonological, and orthographic levels (e.g. Federmeier & Kutas, 1999; DeLong et al., 2005; Van Berkum et al.,

2005; Wicha et al., 2004; Laszlo & Federmeier, 2009). The debate has thus shifted away from simply asking whether prediction can occur to questions about how general of a strategy prediction might be under different processing conditions (cf. Kutas, DeLong, & Smith, 2011).

Federmeier and colleagues have suggested that the young adult brain may routinely engage more than one comprehension strategy by distributing functions across the two cerebral hemispheres, with the left hemisphere more likely to use a prediction-based approach (see Federmeier, 2007). Responses during normal comprehension thus summate across multiple mechanisms unfolding in parallel (Wlotko & Federmeier, 2007). However, the tendency to use predictive processing strategies has also been shown to vary across processing circumstances, especially as a factor of normal aging. Older adults as a group do not seem to engage predictive mechanisms to the same extent as young adults: unlike younger adults, they show no evidence of pre-activating semantic features of predictable words (Federmeier et al., 2002), nor do they seem to show processing consequences associated with encountering unpredictable words (e.g. Federmeier, Kutas, & Schul, 2010). Therefore, we may expect to find a different pattern characterizing ERP effects as a function of contextual strength for older adults. Indeed, Federmeier & Kutas (2005) recorded ERPs as older adults read plausible sentences that were either strongly or weakly constraining for the sentence-final critical words. As for younger adults, N400 amplitudes were reduced in magnitude (facilitated) for strongly compared to weakly constraining sentences. However, the effect of constraint was reduced (and delayed) for older compared to younger adults. These results show that even in simple sentences, older adults gain less from contextual information, and are delayed in the use of the contextual information they can exploit.

Because the evenly graded response in the young adult brain seems to arise from the combination of more predictive and more integrative modes of processing, divided between the two hemispheres (Wlotko & Federmeier, 2007), and because older adults make less use of anticipatory comprehension mechanisms, it is possible that older adults' N400 responses would not show an evenly graded pattern as a function of contextual constraint. No study has previously reported N400 responses from older adults across the full range of cloze probability, motivating the present experiment, which makes use of sentences that parametrically vary across the entire scope of expectancy, from 0 to 100%. To examine the effects of contextual strength in fine detail, we look at N400 responses with statistical tools (item-level ERP regression analysis and multilevel linear modeling) that take into account variation at both the subject and item level, for both younger and older adults. We expect young adult N400s to pattern with cloze probability in an inverse linear manner, as prior studies have shown at the grand-average level. For older adults, we expect smaller N400 effects overall (cf. Federmeier & Kutas, 2005; Federmeier et al., 2002, 2003). Of particular interest is to characterize whether age-related changes in comprehension are similar across the full range of constraint, or whether changes manifest differently in strongly versus weakly constraining contexts, perhaps owing to shifts in modes of processing engaged across age.

In another report (Wlotko & Federmeier, submitted), we have also used an items-based approach in young adults to examine context effects that emerge after the N400 --

specifically, a left frontal negativity (600–900 ms post-stimulus-onset) that was related to the number of unique completions generated by participants in cloze probability norming. In moderately strong sentence contexts, the magnitude of the negativity was greater for frames that elicited fewer unique completions in the cloze probability norming. We hypothesized that the effect could be similar to "frame-shifting" as described by Coulson (e.g., Coulson & Kutas, 2001), a phenomenon that occurs when a reconsideration of prior context is prompted by analyzing current input. This type of process could be engaged when participants consider a likely alternative completion to a moderately constraining sentence context, rather than the actual ending encountered in the experiment. Although a number of studies have examined related phenomena in young adults (e.g., jokes: Coulson et al. 2001, 2005; referential processing: Van Berkum et al., 2007), this type of effect has not been characterized for older adults. Here, therefore, we also examine the late negativity elicited by older adults to determine whether it is sensitive to context in the same manner as for young adults.

Method

Participants

Two groups of native English speakers were recruited from the University of Illinois and surrounding community and compensated with course credit or cash payment for their time. The sixteen young adults (8 female, 8 male) between ages 18 and 27 (mean age 19.7 years) and twenty older adults (9 female, 11 male) between ages 60 and 80 (mean age 67.8 years) were screened for normal or corrected-to-normal vision, no history of neurological or psychiatric disorders, no major head trauma, and no neuroactive medication in the preceding 6 months. All older adults reported attaining at minimum a high school diploma or equivalent; all but five of these participants reported some amount of post-secondary education, and six completed an advanced degree. Thus, as a group, the older adults were more educated than the sample of young adults and were recruited from the same university community as the young adults, mitigating the possibility that the age-related changes uncovered in this study were driven solely by educational and other socioeconomic factors. In addition, the Mini-Mental State Exam was administered to the older adults in order to screen for the possibility of dementia, and all participants completing the study had a score greater than 25 (mean 27.95, median 28).

Participants in both groups were right-handed as assessed by self-report and by the Edinburgh Handedness Inventory (Oldfield, 1971) with a mean handedness quotient of .74 for the young adults and .83 for the older adults. Data from two additional older participants were excluded from analysis due to high levels of artifactual contamination of EEG records. Written informed consent was obtained from all participants, and experimental protocols were approved by the Institutional Review Board at the University of Illinois at Urbana-Champaign.

Materials

Three hundred sentences that varied continuously in the level of constraint were used in this study. Constraint was defined on the basis of cloze probability norming as the highest cloze

completion for a particular sentence. Sentences were selected from a larger set that had extensive cloze norming data, collected from young adults at the University of California, San Diego (more details of the norming procedures can be found in Federmeier et al., 2007 and Wlotko & Federmeier, 2007; the current set of materials contains about 75% overlap with those prior studies). The cloze procedure used for this norming directed participants to give the word that “first comes to mind” given the sentence frame, as well as two additional sentence completions for each frame. This procedure allows the norming data to potentially include more information about words that are not often generated in the cloze task, such as more unexpected endings.

Sentences sampled the full range of constraint, including weakly constraining sentences paired with wholly unexpected endings (cloze probability near zero) and strongly constraining sentences with their best completions (cloze probability near 100%). Fifty sentences containing wholly unexpected (although still plausible) completions were chosen from the materials constructed for the studies reported in Federmeier et al., 2007 and Wlotko & Federmeier, 2007. These were always weakly constraining (i.e. the cloze value for the word with the highest close probability for that frame did not exceed 42%), but the sentence-final words were never the word with the highest cloze probability for that sentence frame. The actual cloze value (totaling the first and secondary cloze values) for the unexpected words was near zero and did not exceed 10%. For the remaining 250 sentences, sentence final words were always the “best completion” (the word with the highest close probability for that frame), meaning that effects of constraint and expectancy are not separable in these materials. We evenly divided the sentences into 5 cloze probability ranges with 50 sentences each: 10–30%, 30–50%, 50–75%, 75–90%, and 90–100%. Representative examples of the stimulus materials are presented in Table 1.

The sentence-final critical words did not differ across these cloze ranges in lexical characteristics of word frequency, length, concreteness, imageability, familiarity, or word class (all retrieved from the MRC Psycholinguistic Database; Wilson, 1988). A summary of these characteristics is presented in Table 2. Because the characteristics were relatively homogenous across the stimulus set, other groupings of cloze did not cause major differences in lexical characteristics. In addition, cloze probability is not correlated with any of the lexical characteristics.

Two pseudo-random orders of sentences were generated, and half of the participants of each gender were presented with one order and half with the other. Because some critical words completed more than one sentence frame in the set, the conditions for randomization precluded presentation of any critical word within thirty sentences of another occurrence of that critical word.

Expectancy and plausibility judgments—Each sentence in the stimulus set was rated for expectancy and plausibility by 19–21 young adults and 7–9 older adults at the University of Illinois. None of these participants were recruited for the ERP study. The experimental sentences were intermingled with stimuli from other experiments and appeared in several different lists. The lists were also padded with anomalous sentences in order to encourage a full use of the rating scale from completely expected and plausible to entirely unexpected

and implausible. Participants were asked to rate, on a scale of 1–7, how expected each sentence ending was, and in a separate judgment, how much the ending “makes sense”. A rating of 1 indicated very expected or entirely sensible, and a rating of 7 indicated totally unexpected or completely nonsensical. Examples were given to attempt to draw participants' attention to the difference between expectancy and plausibility.

For the sentences comprising the ERP study herein, Table 3 shows the expectancy and plausibility ratings for each cloze bin described above, for younger and older adults. The Pearson correlation coefficient between the two sets of ratings is .86 for expectancy and .81 for plausibility. Thus, younger and older adults were similar in their offline judgments of contextual information in sentences. Because of the larger number of young participants who rated the stimuli, leading to more robust norming data, analyses focused on predicting both younger and older adults' brain activity with the ratings from the young participants, but no conclusions were altered when the older adults' ratings were used instead.

Procedure

After electrode application, participants were seated in an electrically shielded chamber or a dimly lit room 100 cm in front of a 21" CRT computer monitor. Each trial began with a warning sign (four pluses in the center of the screen) presented for 500 ms. The duration of the blank screen between the warning sign and the first word of each sentence varied randomly from 500 to 1200 ms to avoid averaging in slow potentials associated with anticipation of sentence onset. Sentences were presented word-by-word for 200 ms with a 300 ms ISI in a Helvetica 22-point font with black text on a white background. A 2.5 second blank screen followed each sentence-final word. Each trial ended with the prompt "Please go on when you are ready." Participants initiated the next trial with a button press.

Participants were asked to minimize blinks, eye movements, and muscle activity while reading. The pause between each trial allowed participants to self-pace the experiment to some extent, but participants were encouraged to move reasonably quickly through the sentences so as not to extensively lengthen the recording portion of the experiment, which took about 1 hour on average. In addition to the pause between each trial, an experimenter-controlled break was given after every 30 sentences, during which the experimenter conversed with the participant to ensure comfort and alertness.

After the recording session, participants completed a recognition test consisting of a list of 60 sentences, each missing its final word. Half of the sentences were read during the experiment and half were taken from various cloze probability norming sets and were not seen during the experiment. Participants were asked to mark each sentence context that they remembered reading during the experiment, and for those, to also fill in the missing word that was read during the experiment. Of the sentences seen during the experiment, 5 came from each 50-trial cloze range. Within these 5 sentences, the average cloze value corresponded to the mean of the experiment-wide cloze range, and word frequency did not differ among these conditions within the recognition test. Participants also completed a short set of neuropsychological measures including verbal fluency (category and letter) and author and magazine recognition questionnaires (Stanovich & West, 1989), and for older adults only, a reading span test (Daneman and Carpenter, 1980). Examination of these measures in

relation to ERP results did not uncover meaningful patterns of correlations for the findings reported here.

EEG recording and processing—EEG was recorded from twenty-six geodesically arranged sites on the scalp using silver-silver chloride electrodes embedded in an Electro-cap. The sites, shown in Figure 1, are Midline Prefrontal (MiPf), Left and Right Medial Prefrontal (LMPf and RMPf), Left and Right Lateral Prefrontal (LLPf and RLPf), Left and Right Medial Frontal (LMFr and RMFr), Left and Right Mediolateral Frontal (LDFr and RDFr), Left and Right Lateral Frontal (LLFr and RLFr), Midline Central (MiCe), Left and Right Medial Central (LMCe and RMCe), Left and Right Mediolateral Central (LDCe and RDCe), Midline Parietal (MiPa), Left and Right Mediolateral Parietal (LDPa and RDPa), Left and Right Lateral Temporal (LLTe and RLTe), Midline Occipital (MiOc), Left and Right Medial Occipital (LMOc and RMOc), and Left and Right Lateral Occipital (LLOc and RLOc). The position of the cap was determined by placing the Midline Prefrontal electrode (MiPf) at 10% of the nasion-inion distance from the nasion, the Midline Occipital (MiOc) electrode at approximately 10% of the distance from the inion, and the Midline Central (MiCe) electrode halfway between the mastoid processes. The electrodes were referenced online to the left mastoid and later referenced offline to the average of the left and right mastoids. Eye movements were monitored using a bipolar recording of EOG with electrodes placed on the outer canthus of each eye. Blinks were monitored with an electrode placed over the infraorbital ridge of the left eye, referenced to the left mastoid. Electrode impedances were kept below 4 k Ω and signals were amplified with Sensorium amplifiers set at a bandpass of 0.02 to 100 Hz. EEG was sampled at 250 Hz and saved on a hard drive.

EEG records were examined and marked for EOG, EMG, or other artifactual contamination. Artifactual trials containing eye blinks were corrected (see Dale, 1994 for the procedure) and added back into the EEG record, except for 8 young and 8 older participants with too few or too many blink artifacts for effective correction. The remaining artifactual trials (13% for young adults and 16% for older adults) were excluded from further analysis.

The data were rereferenced to the algebraic mean of the left and right mastoids, and averages of artifact-free ERPs were calculated for critical words, computed from 100 ms before the onset of critical words to 920 ms after, with subtraction of the 100 millisecond pre-stimulus baseline. Measurements for mean amplitudes were taken after a digital bandpass filter of .2 to 20 Hz was applied; for peak latencies, a bandpass filter of .2 to 10 Hz was used.

Results

Recognition Test

Young adults correctly recognized an average of 19.1 sentences from the experiment (64%) and false alarmed to an average of 1.3 unseen items on the recognition test (mean $d' = 2.14$). Of the correctly recognized sentences, the younger adults correctly completed an average of 15.3 missing sentence-final words from the experiment. Whereas older adults' performance on correctly recognized items (18.3 items, 61%) and correctly completed items (13.7) was similar to that of young adults, older adults were more likely to incorrectly endorse unseen

items on the test (mean false alarms = 7.1 items), leading to a significant difference in d' across the two groups (older mean $d' = 1.23$; $F(1,34)=20.16$, $p < .001$). These results indicate that both groups of participants were attending to the experimental material, while also showing the often reported effect of increased false alarm rates in memory tasks for older populations (see, e.g., Benjamin, 2001).

ERPs

Between 0 and 250 ms post-stimulus onset, both age groups show the characteristic series of evoked potentials elicited by visual stimuli, as displayed in Figure 1. Consistent with past studies of word-by-word reading examining sentence-final critical words, visual inspection reveals that posterior potentials are slightly reduced in size for older adults, whereas, anteriorly, the N1 is much larger in older compared to younger adults and the P2 is notably reduced (King & Kutas, 1995; Wlotko et al., 2010). ERPs are overlapped for the two age groups at two frontal, central, and occipital sites in Figure 2.

N400—Following the sensory components, a broadly distributed negative-going wave, the N400, peaks at 381 ms for young adults and 393 ms for older adults (when assessed as peak latency for the three largest N400 conditions, between 300–500 ms post-stimulus onset at the three medial parietal sites where N400s are largest). This delay is not significant [$F(1,34)=1.4$, $p=.2446$], although this measurement is likely to be susceptible to overall group differences in waveform morphology (see Figure 3). Of greater interest is the typical N400 sentence-level expectancy effect, assessed by first computing a difference wave via a point-by-point subtraction of the ERP for strongly expected endings from the ERP for unexpected endings. Peak latency of this difference was measured between 300–500 ms at the same three parietal sites. Mean amplitudes were then computed for the two conditions (strongly expected and unexpected) by averaging over a 200 ms window around the peak of the difference for each age group, over all electrode sites (280–480 ms for young adults; 325–525 ms for older adults). The effect of expectancy on N400 mean amplitude was significant for both groups, showing that the ability to gain facilitation from sentence context is retained with age (for young, $F(1,15)=19.18$, $p=.0005$, interaction with Electrode, $F(25,375)=9.19$, $p<.0001$, $\epsilon=.1969$ ¹; for older, $F(1,19)=25.48$, $p=.0001$, interaction with Electrode, $F(25,475)=17.32$, $p=.0032$, $\epsilon=.1492$). However, consistent with past work, the effect is significantly smaller (2.2 μV vs 1.1 μV ; $F(1,34)=4.47$, $p=.0419$) and delayed by about 45 ms (382 ms vs 425 ms; $F(1,34)=11.52$, $p=.0018$) for older adults, attesting to their less efficient use of context.

Regressions of N400 amplitude on cloze probability - young adults: In another report (Wlotko & Federmeier, submitted), we established the group-level correlation between N400 amplitude and cloze probability for the group of younger adults. We found that correlation coefficients ranged from .3 over the front of the head to .9 over right posterior regions, demonstrating that the pattern of correlation with cloze probability matches the well-known N400 distribution over the scalp (cf. Delong et al., 2005) and validating our

¹For repeated measures ANOVA, p-values for all tests with more than one degree of freedom in the numerator are adjusted with the Huynh-Feldt correction for sphericity and the epsilon value is provided.

materials for the investigation of sentential context across a wide range of contextual strength.

This type of analysis, while informative, disregards subject variability. In order to show that results are consistent across subjects, N400 amplitudes were analyzed with a hierarchical linear model, allowing the effect of cloze to be random across subjects rather than fixed. Thus, subject variability and inter-subject dependency is taken into account. This analysis allows us to determine the consistency of patterns across subjects through estimates of the random effects, provides statistically efficient estimates of the fixed effects, and further, by explicitly modeling within-subject dependency, produces correct standard errors for the fixed effects estimates (see, e.g., Goldstein, 2011). N400s for 12 cloze bins (each of the six 50-item cloze bins divided in half) were measured from 280–480 ms at the 13 electrode sites with correlations at the group level of .8 or greater². These sites (LMCe, RMCE, MiCe, RDCe, RLTe, LDPa, RDPa, MiPa, LMOc, RMOC, LLOc, RLOc, MiOc) are all contiguous in the posterior region, consistent with N400 distribution. The null model (with no predictor) naturally gives an estimate of the between-subject variability, leading to calculation of the intra-class correlation, which for this dataset is .36³. This statistic can be interpreted as showing that 36% of the variability in N400 amplitude is between subjects, reflecting the relatively large individual differences in ERPs across subjects and suggesting the appropriateness of the hierarchical linear model for this analysis. When the model with cloze (parametrized as the average cloze value in each of the 12 bins) as a predictor is fit, the estimated regression coefficient is 3.39 [$p < .0001$, using a t distribution with 15 degrees of freedom; 95% confidence interval 2.43, 4.35], with variability of 2.97. Thus individual subject effects range from 1.6 to 7.6, showing consistency in direction of the effect across subjects (see figure A1 for the estimated random effects). These estimates can be interpreted as the predicted size of the N400 effect for the full range of cloze probability, as cloze probability was coded as a proportion ranging from 0 to 1.

Although the hierarchical model naturally considers inter-subject variability by grouping data points within subjects, another way to examine the relationship between an electrophysiological response and some stimulus variable is at the item level⁴. This analysis takes full advantage of the continuous nature of the manipulation of cloze probability in this study. Average N400 amplitudes for each item (across subjects) were measured between 280–480 ms post-stimulus onset at the same thirteen electrode sites with group-level correlations above .8. Figure 4 displays a scatter plot of cloze probability versus the single-item N400 amplitudes across these sites. The correlation between N400 and cloze probability at the item level is .42 ($p < .0001$, confidence interval .32 to .51). This is, of course, a smaller correlation than at the group level because the correlation coefficient takes into account both the strength of relationship between the two variables as well as the

²Note that it is possible to construct a three-level model which also includes electrode site as a level within subjects; however, here we chose to focus on a set of sites known to produce well-studied N400 effects, averaging over the spatial dimension.

³See appendix for details of model fitting procedure.

⁴This analysis could also be done with a hierarchical linear model. However, given that the estimates of the fixed effects will be statistically consistent, and intersubject variability is not the focus of this analysis, we chose to conduct our analysis with ordinary least-squares regression and correlation in order to take advantage of the closed-form solution. This not only reduces computational demands, but also allows the more direct evaluation of R^2 measures in order to examine sensitivity to contextual variables for the two age groups.

variability (deviation from the linear function relating the two variables). In conducting the ordinary least-squares regression, we find that the regression coefficients are quite similar (3.14 for the binned analysis vs 3.16 for the items analysis). Thus, at both the group and the item level, one can expect about a 3 microvolt effect of cloze probability in plausible sentences that range from very strongly expected to wholly unexpected, and, consequently, a one percent increase in cloze corresponds to a decrease in N400 amplitude of about .03 microvolts.

To ensure that the item-level analysis provides distributional information similar to the group level, the correlation at the item level was assessed at each electrode. The interpolated correlation values are mapped in Figure 5. Like the group-level analysis, item correlations create a right-posterior distribution over the scalp.

Contextual strength and other sentence-level variables: Cloze probability explains 90% of variance in N400 amplitude in group-level ERPs and 17% at the item level in this dataset⁵. In this investigation, we were interested in taking advantage of the item-level analysis to determine whether some other sentential variables also affect N400 amplitude over and above the effect of cloze probability. In addition, the wide range of contextual constraint employed in this study allowed a more detailed look at how N400 amplitudes track variation in contextual information than has previously been considered. Norming data for the materials included posthoc ratings of expectancy and plausibility. Both of these variables are highly correlated with cloze probability, but their precise relationship varies by cloze range. Because our set contains 50 items in the 90–100 cloze range and 50 unexpected items, we removed these two sets of items from our dataset in order to restrict our analysis to a more uniform distribution of cloze. In this range, 8% of variance is accounted for by cloze probability when assessed at the item level. Across this range, expectancy and plausibility ratings do not explain more variance when added to a regression model. We next divided the cloze range in half and found that the relationship between cloze probability and N400 is stronger at lower levels of cloze: whereas 7% of variance is explained in the 10–50% range, less than 1% is explained in the 50–90% range (and is not significant). Table 4 displays results of these regression analyses.

We also examined the 90–100% range and unexpected endings separately. No variables significantly explain N400 amplitude in the 90–100% range, likely due to the restricted range. Cloze probability was also not significantly related to N400 amplitude for the unexpected items, again likely due to the small range of cloze values (mostly zero)⁶. The unexpected items in the stimulus set were by definition never produced very often in the cloze probability norming. Thus, these items may be particularly subject to sparse sampling in the cloze data, even for our cloze procedure that included data on secondary completions. Expectancy and plausibility ratings, therefore, may be particularly likely to track N400 amplitudes for the unexpected items. Indeed, both plausibility and expectancy ratings

⁵In all analyses reported here, the “total” cloze value was used – that is, proportion of participants completing each sentence frame with a particular word, including words generated as a secondary completion. In the regression models presented here, including secondary cloze as its own variable nearly always explained additional variance in N400 amplitude over and above cloze probability as a first completion, attesting to the usefulness of the additional information contained in this expanded cloze measure.

⁶Even with this restricted range, 6% of variance is explained by cloze alone, and the p-value is .07.

explained significant variance in N400 amplitude for unexpected items (expectancy $R^2=.11$, $p=.02$, plausibility $R^2=.1$, $p=.03$), showing that the N400 is sensitive to small variations in contextual information available in the contexts that is not apparent in cloze values alone.

N400 regressions - older adults: Group-level correlations of average cloze probability with N400 amplitude (measured for the twelve 25-item bins, between 325 and 525 ms post-stimulus onset) range from .93 at the middle occipital site to a minimum of $-.37$ at the left lateral prefrontal site. Correlations are consistent with the typical N400 scalp distribution, as they decrease anteriorly and to the left. Different from young adults, the sign of the correlations reverses over left frontal sites; this is likely because N400s for older adults are smaller and later than for young adults and are thus more likely to be contaminated by other effects (e.g. the left frontal negativity discussed below).

A linear mixed model with no predictors fit to the N400 data of older adults reveals an intra-class correlation (again, providing a measure of between-subjects variability) of .42, similar to young adults. For the older adults, a model allowing a random effect of cloze leads to an estimated fixed effect of 2.18 ($p<.0001$, using a t distribution with 19 degrees of freedom; 95% confidence interval 1.61, 2.74), with variability estimated at 1.32. Individual subject effects range from $-.8$ to 3.81, thus showing that overall, young adults and older adults follow a similar relationship between sentential context and N400 amplitude. However, consistent with the smaller overall N400 effect for older adults, the estimated fixed effect is smaller (e.g. the coefficient represents the predicted change in N400 amplitude from a wholly unexpected item with 0% cloze to a completely expected item with 100% cloze; this value is predicted to be smaller for older adults). Additionally, estimated random effects for all young adults exceeded 1.5, through a maximum of 7.6, but no older participant had an effect larger than 4.0, and one older participant's effect was negative (see Figure A1). In accord with the age differences suggested by these separate models, a direct comparison of the between-group effect of cloze probability revealed a significant Group x Cloze interaction [$F(1,34)=5.84$, $p=0.021$]; see Appendix for more details.

In turning to the item-level analysis, we find that item N400 amplitudes correlate with cloze probability, giving a Pearson coefficient of .42 ($p<.0001$, 95% confidence interval [.32, .51], $R^2=.18$) when assessed at the 13 sites for which group-level correlations are greater than .8 (the same sites as for young adults), very similar to the young adult correlation. Item-level correlations were computed at all 26 scalp sites and interpolated values are presented in Figure 5. The correlation map is comparable to that of the young adults, showing that overall, young and older adults N400s behave similarly.

When truncating the stimulus set to remove the 50 most expected and 50 unexpected stimuli to create a more uniform distribution, R^2 becomes .12 ($p<.0001$). However, after dividing the sentences into the 50–90% and 10–50% cloze ranges described for young adults, in neither case is cloze probability significantly associated with N400 amplitude (50–90%, $R^2=0$, $p=.99$; 10–50%, $R^2=.01$, $p=.32$). In no ranges do we find that ratings of expectancy or plausibility add significantly more variance than cloze probability alone. Thus, the differences in the patterns for younger compared to older adults are not simply because the

older adult brain is less sensitive during online comprehension to the kind of information captured by cloze probability.

N400 regressions – group comparison: Taken together, the results of our item analyses (taking full advantage of the continuous nature of our design) suggest that variation in N400 amplitude for older adults is less sensitive to the amount of contextual information available in sentential contexts compared to the young adult brain wave pattern. In addition to the overall smaller and delayed effect of message-level expectancy, the older adult brain gains less facilitation from weak contextual information. To directly compare brain wave patterns from the younger and older adults in this study, the hierarchical linear model described above was re-fit separately for strong and weak constraint conditions in the ERPs. Based on the patterns of item-level correlations above, age differences in sensitivity to contextual information as a function of age should emerge for weak but not strong contexts. All parameters from these models are presented in Table A2. The theoretically important results are from the Group x Cloze interaction terms, significant for weak contexts [$F(1,34)=7.88$, $p=.0082$], but not for strong contexts [$F(1,34)=0.0$, $p=.995$]. Thus, these results support the suggestion from the item-level analyses that older adults show less graded variation in N400 amplitude related to sentence-level constraint.

Late left-lateralized frontal negativity—In prior work using the data collected from the young subjects, we described a left-lateralized frontal negativity that emerged 400–500 ms post-stimulus onset (Wlotko & Federmeier, submitted). This potential was largest for items in the 75–90% cloze range, and within this cloze range, the magnitude of negativity at the item level inversely correlated with the number of possible completions generated by participants in the cloze norming task (that is, the negativity was larger for items with fewer alternatives). The effect was taken to reflect reinterpretation of the context information in cases wherein some participants developed an alternate reading of the sentence, different from the one entailed by the presented best completion (as indexed by a small number of alternative endings provided in the cloze norming, signaling a likely alternate completion for the sentence).

The older group also generated a similar effect, in an even broader range of cloze probabilities, which extended into the 90–100% range. An omnibus ANOVA of mean amplitudes in the 600–900 ms time window revealed a main effect of Condition [$F(5,95)=6.30$, $p<.0001$, $\epsilon=1.000$] and a Condition x Electrode interaction [$F(125,2375)=5.60$, $p<.0001$, $\epsilon=.1055$]. For the older adults, both the 90–100% and 75–90% cloze ranges were significantly different from all other bins when measured 600–900 ms post-stimulus onset at three left anterior sites at the focus of the effect [all Fs, with 1 and 19 degrees of freedom, are greater than 7.42, leading to all $p<.013$], but these two conditions did not significantly differ from one another [$F(1,19)=.98$, $p=.33$]. To directly substantiate the apparent difference in range of constraint that elicits the negative-going effect across the two groups, difference waves for the 90–100% bin vs Unexpected items and the 75–100% bin vs Unexpected items were measured over 600–900 ms at the same three channels and subjected to a between-groups ANOVA. While the negativity was significantly larger for older adults in the 90–100% range [$F(1,34)=5.91$, $p=0.021$], there was no significant

difference for the effect in the 75–90% range, where both groups evidenced the negativity [$F(1,34)=0.08$, $p=.777$].

To explore the sensitivity of the effect to the number of unique completions from the norming results, we analyzed the data at the item level as we previously did for young adults in Wlotko & Federmeier (submitted). The magnitude of the negativity (relative to baseline) for older adults was significantly related to the number of alternative endings generated in the norming study, as it was for young adults, but again over a broader range of cloze values. While not a significant correlation when restricted to the cloze range for which the young adults demonstrate the effect (75–90% cloze, $r=.19$, $p=.18$), the effect is significant for older adults if the cloze range is expanded to 50–100% ($r=.18$, $p=.03$)⁷. Thus, the effect seems functionally similar for younger and older adults, but over a larger range of strength of sentential context for older adults. The item mean amplitudes for the negativity versus number of unique alternate endings is displayed in Figure 6, overplotted with the data from the young adults.

Discussion

To build the database of knowledge about cognitive and neural mechanisms of language comprehension, and how they may change over the lifespan, researchers have drawn upon a wide variety of paradigms and methodologies. Event-related potential studies have made a particularly substantive contribution, in providing direct, specific, and multidimensional measures of neural activity with a degree of temporal resolution necessary to investigate comprehension in real-time -- a combination of characteristics offered by few other commonly used methodologies. Studies using the N400 have shown that context information accrues over the course of a sentence or larger discourse and that the availability of this context information can have a profound influence on word processing. Shortly after its discovery, the close relationship between N400 amplitude and the fit of the N400-eliciting word into its context (independent of the constraint of the context itself) was observed for young adults (Kutas & Hillyard, 1984). This fact has been exploited by a large number of studies to examine the architecture of the comprehension system from a number of vantage points (reviewed in Kutas & Federmeier, 2011) and has led to the conceptualization of the N400 as a window into initial semantic processing elicited by a potentially meaningful stimulus.

In this study, we extend these important, classic findings. Using a set of entirely plausible sentences that contain a continuous variation of cloze probability over the full range of

⁷The 90–100% cloze bin includes some sentences for which no alternate endings were generated in the cloze norming (i.e. cloze probability is 100%). For the analysis presented here, items with 0 or 1 alternate endings were removed, on the idea that these items may not be particularly amenable to any reinterpretation, even for older adults (this totaled 2 items in the stimulus set with no alternative endings and 12 items with one alternate possibility, each generally given by only one participant in the norming; all but one of these items came from the 90–100% cloze range, and most of them came from above 95% cloze, meaning that with this criterion applied, the average cloze value for the 90–100% bin became 93% with a maximum of 97%). With these items included, the correlation was not significant ($r=.12$, $p=.16$). Further, the 50–100% range is the widest range where older adults evidence a significant correlation, although the pattern remains the same and is significant if the range is restricted to the 60–100% range. Thus, the correlation for older adults is significant over a wider range of constraint than clearly elicits a large negativity, but behaves as would be expected if related to reconsideration of contextual material, in that it is less robust for weak constraint as well as for especially strongly constraining contexts without possible alternative completions.

possible values, we conducted parametric analyses of several types to provide stronger statistical support for the long-standing claims about the relationship between N400 amplitude and cloze probability and also several other sentence-level variables. Moreover, we tested a sample of older adults to determine in a fine-grained manner how the use of sentential context changes over the lifespan -- a question that was previously unexplored in the ERP literature at this level of detail.

At the grand average level, we found that for young adults, the N400 has a strong relationship with cloze probability, and correlation maps for this sample matched the well-known N400 scalp distribution (see also Wlotko & Federmeier, submitted), corroborating numerous other studies, including the classic Kutas & Hillyard (1984) report. Our examination of the data with a multi-level model, which explicitly takes into account inter-subject variability and intra-subject dependence, provided additional statistical evidence that the phenomenon observed at the grand average level is concordant with patterns elicited by all individual subjects.

We also examined ERPs at the item level. We again found the same relationship between N400 amplitude and the availability of supportive contextual information and the same distribution over the scalp. The item-level analysis revealed that it is possible to explain 17% of variation in our N400 data with cloze probability alone and that significant variance could still be explained when the tails of the cloze distribution were removed (i.e. the correlations were not entirely driven by either the most strongly expected or completely unexpected items in the set). Although the item-level correlations are smaller than for the grand average, substantial between-subject variability (the intra-class correlation was 36% in our sample of young adults), combined with a relatively small sample size (typical of ERP sentence processing studies), would be expected to produce smaller estimates of variance explained. Importantly, however, this finding demonstrates that when looking for robust effects of parametric variables, patterns can be uncovered without employing the large number of subjects that would be necessary under other circumstances (cf. Laszlo & Federmeier, 2011).

After dividing the range in half (10–50% and 50–90%), we found that the variation in N400 amplitude is stronger at lower levels of cloze probability. Thus, with higher levels of contextual information, there may be more individual differences in what contextual information is used and how much facilitation in processing can be gained from the context. Indeed, the use of context information can affect multiple aspects of processing, as revealed by the left frontal negativity for moderately constraining contexts. As the onset of this negativity overlaps in time and to some extent in distribution with the N400, the reduced sensitivity of our measurements to cloze probability in the higher levels of cloze could also be due in part to component overlap.

Operationally defining the amount or quality of contextual information by cloze probability has worked extraordinarily well in psycholinguistics, even though the measure is not particularly specific or informative about exactly what kind of information is available in the context. Further, necessarily sparse sampling in cloze norming studies means that there is often not much information at all about weak contexts or words that are not often generated

by participants in norming studies. In order to gain additional information about our experimental stimuli, we collected post-hoc judgments of expectancy and plausibility for all of our sentence endings (from a different group of participants than those in the ERP studies). Across the entire range, these ratings are highly correlated with cloze probability, which would be expected because they all measure at some level the amount of contextual information in a sentence context in relation to a sentence-final word.

However, different processes and information may contribute to an online brain response during comprehension versus an offline response, and, even within offline tasks, to the generation of an ending (as during a cloze task) versus an expectancy judgment. Our subjective ratings allowed us to ask what information in these offline judgments, over what levels of contextual constraint, could be used to predict brain activity during online comprehension. Of particular interest was whether subjective ratings could be useful in explaining variation in N400 for items with sparse sampling in the cloze norming task. Indeed, the correlation between cloze probability and these other ratings diminishes at levels of weak constraint. In fact, both expectancy and plausibility ratings explained significant variance for the unexpected endings when cloze probability did not. This supports the idea that subjective post-hoc ratings provide more information than cloze probability itself for unexpected endings, possibly because of the necessarily limited information about unexpected endings that cloze probability provides. It further demonstrates the reliable sensitivity of the N400 to small amounts of contextual information, and makes the point that researchers employing sentences with weakly constrained or unexpected endings in their designs should take care to not only match cloze probability but to also consider other measures of contextual fit across conditions.

Across these analyses, we characterized, in a more precise manner than has been done before, younger adults' responses to sentential context measured in a variety of ways. To examine how the sensitivity to contextual information in comprehension may change with healthy aging, we also asked a group of older adults to read the same materials as we recorded ERPs. Prior work has shown that older adults seem to rely on different sets of strategies or resources for sentence comprehension (see Wlotko et al., 2010). In particular, older adults are less likely to use predictive processing mechanisms (Federmeier, Kutas, & Schul, 2010) and, perhaps related to this, have been found to make poorer use of context information to shape semantic processing as indexed by the N400 (e.g., Federmeier et al., 2002). Differences in the use of predictive versus more bottom-up (integrative) comprehension strategies across the two cerebral hemispheres also result in different patterns of responses to variations in the availability of context information (Wlotko & Federmeier, 2007). Indeed, the evenly graded pattern of N400 response to cloze probability described here and previously for young adults seems to result from the combined contributions of the mechanisms instantiated in the two hemispheres, neither of which alone shows the canonical pattern. Thus, one would predict that differences in the type or diversity of processing strategies employed by older adults would result in changes in the pattern of response to graded cloze probability -- something that has not, to our knowledge, previously been examined at this level of detail.

Here we used a set of entirely normal, plausible sentences, for which participants had no task other than to pay attention to and comprehend the sentences. As older adults may be more sensitive to task demands, this homogenous grouping of sentences, with no anomalies, in a very simple paradigm for which participants would likely not invoke any special strategies, allows an investigation of the use of context during more "normal" comprehension across a wide range of contextual strength.

Overall, as expected, older adults' N400 responses were modulated by the availability of context information, as amplitudes decreased with increased constraint, and the correlation pattern was qualitatively similar to that of young adults and had a very similar scalp distribution. As in past studies, however, the effects of context were smaller and delayed for older adults. Further, the fine grained examination showed that the effect of sentence context was not as strong for older adults at either high or low levels of constraint. Older adults, like young adults, have large stable N400s for unexpected endings, which are slightly facilitated for weak context (particularly in the later time window, which may represent late positive complex (LPC) activity), and then exhibit more facilitation when contextual information reaches about 50%. Thus, instead of the acute sensitivity of N400 amplitude to variation in contextual information seen for the young adults, the responses of older adults seem to essentially bifurcate the context space, such that they either gain some facilitation from the context (albeit to a lesser degree overall than young), or not. Further, older adults' reduced sensitivity to contextual information extended to decreased responsiveness of N400 amplitudes to variation in post-hoc judgments compared to young adults.

The patterns of N400 activity observed for older adults are thus distinct from the typical pattern of young adults. However, young adults also sometimes fail to show the evenly graded N400 response to variation of contextual information, particularly when processing is biased to one cerebral hemisphere over the other. We have previously shown (Wlotko & Federmeier, 2007) that weakly constrained words can be facilitated to the same degree as strongly constrained ones (when processing is biased to the left hemisphere) or not facilitated compared to wholly unexpected items (when processing is biased to the right hemisphere). The graded pattern -- in which weakly constrained words elicit intermediate levels of N400 activity -- thus seems to require the parallel contributions from multiple processing mechanisms across the two hemispheres.

The cloze probability range in which the hemispheres are most different from each other is the same the range wherein older adults show less robust facilitation in the current study (when constraint is weak). A resemblance between patterns of right-hemisphere-biased brain responses in the young and the patterns seen in older adults has been observed across several different paradigms examining sentence processing (Federmeier et al., 2002; see also Wlotko et al., 2008). Therefore, given that the evidence suggests that younger adults may rely on parallel contributions from both hemispheres to effect normal comprehension, we can conjecture that older adults may shift the balance of hemispheric contributions for language processing. Although still a speculative link, other work shows that older adults rely on a different configuration of hemispheric activity compared to young adults for some language tasks (Meyer & Federmeier, 2010). Further work will be needed to uncover the direct source of the decrease in graded sensitivity to context for older adults in relation to

hemispheric asymmetries and differential hemispheric resources. For example, it is important to determine whether the two hemispheres seem to offer more redundant instead of complimentary patterns, perhaps because degraded or inefficient functional connectivity leads to less availability of top-down processing that is required for an anticipatory, predictive comprehension strategy. Alternatively, decreased availability of neural resources may render a predictive strategy less effective or too costly for older adults, who could then rely more on right hemisphere processing mechanisms to achieve comprehension goals. Either of these possibilities might interact with or compound changes across age in working memory resources (e.g., Reuter-Lorenz et al., 2000), inhibition of irrelevant information (e.g., Zacks & Hasher, 1997), or motivational approaches to language endeavors (Stine-Morrow et al., 2006; see also Wlotko et al., 2010).

Whatever the source of the change with age, it is clear that older adults show reduced effects of context overall and reduced sensitivity to finely graded differences in contextual strength and/or expectancy for particular words. These changes in mechanism can have further, downstream consequences -- and we see one example of this in the current study. In the dataset from young adults we observed a late left frontal negativity specifically for sentences of moderately strong constraint (Wlotko & Federmeier, submitted). Because of the differences in time course and scalp distribution, the effect is separable from the central-parietal N400 in the earlier time window. Indeed, we demonstrated in the prior report that item-level amplitudes of the N400 and the later left frontal negativity were not correlated, suggesting that these brain responses are not functionally yoked. Many other sentence processing studies have observed left frontal/anterior negative-going effects (sometimes called LAN effects) with various time courses. As we discuss in Wlotko & Federmeier (submitted), all of these effects may have in common some underlying component of verbal working memory. For example, manipulation of verbal information due to syntactic or structural constraints often manifests in a left-lateralized, anterior, negative-going brain response (e.g., Kluender & Kutas, 1993). Memory demands can also be invoked by circumstances requiring that multiple representations remain active, either in referentially ambiguous (eliciting the Nref effect; e.g., Van Berkum et al. 2007) or semantically ambiguous contexts (see Lee & Federmeier, 2009 for a description of a similar sustained frontal negativity). All of these situations may also require selection of an appropriate representation and/or inhibition of inappropriate one(s).

Based on the similarities in time course and waveform morphology to negativities observed in response to relatively simple plausible sentences (e.g. Kutas, 1993; Kutas & Federmeier, 2005), we hypothesized that the negativity effect observed here could be comparable to one particular manifestation of this type of left frontal/anterior effect, related to the frame-shifting operation described by Coulson in which contextual information is reinterpreted in light of the current input. That is, we linked this response to reinterpretation of contextual information needed when processing a likely possible alternative completion. In this study, we always presented the "best" (highest cloze) completion. But, for moderately strong contexts, some items will have a secondary likely completion, in contrast to very strong contexts, for which there is only one strongly expected completion, and to weak contexts, for which no particular ending is very likely. In those cases in which there is a likely competitor, some participants may have been led to a reading of the context that supported

the competitor, which could then induce a reconsideration of the contextual material upon encountering the best completion instead. For example, the sentence frame "Each night the campers built a huge..." is most often completed (in cloze norming) by the word "fire". However, a sizable minority of participants instead generate "tent" as a good completion for the frame. For participants who might have been constructing a message level meaning related to building a tent, therefore, encountering the word "fire" instead would likely lead to a reconsideration of the meaning of the sentence. Support for the idea that the negativity is tied to the likelihood of a potential secondary competitor was provided by an items-level analysis correlating the negativity with the number of alternate endings generated by participants in the norming study. Sentences with fewer endings are more likely to be those with a relatively common alternative interpretation, whereas sentences of similar cloze probability with a larger number of possible completions may be those which, for some people, are more like weakly constraining contexts (making frame-shifting-like processes less likely). We found that the size of the left frontal negativity was inversely correlated at the item level with the number of alternative completions, such that the fewer alternative completions that were generated, the larger the size of the negativity.

Older adults also show this effect, albeit in a somewhat different form. They showed a similar negativity for items of moderately strong constraint, but also elicited this same effect for sentences of strong constraint. In addition, across an expanded range of constraint compared to young adults, the items with fewer alternate completions engendered a larger negativity, a pattern qualitatively consistent with that of the young adults. Thus, this ERP response likely reflects similar processes in the two age groups, but the older adults engage these processes over a wider range of contextual strength. We suggest that older adults' decreased tendency to recruit predictive processing mechanisms (which results in a different brainwave pattern, a late prefrontal positive effect⁸; see Federmeier et al., 2007; Federmeier, Kutas, and Schul, 2010) might render the strongly constrained sentences more like those of moderate constraint (concordant with the fact that overall context effects for older adults are smaller). That is, lack of prediction may cause a reading of the context to go in a different direction or create a less robust message-level representation, making it correspondingly more likely that reinterpretation will be engaged, even for strong contexts (which, online, lead to more consistent predictions in young adults). Given that responses in cloze norming tasks are generally consistent across age (e.g. Lahar, Tun, & Wingfield, 2004) and that older adults perform similarly to young adults in offline comprehension tasks (e.g., Light and Burke, 1993), this hypothesis lends support to the idea that neural changes can result in shifts in pathways used to reach similar comprehension outcomes across age.

In summary, then, in this study we have shown strong support for long-standing claims about the language comprehension system and extended our knowledge of comprehension mechanisms in young and older adults with a fine grained examination of brain responses to

⁸A number of considerations argue against the fact that the effect observed here is better characterized as this type of positive-polarity effect for the lower constraint conditions rather than a negative-going wave for the strong constraint conditions. First, our design does not include any violation of constraint; that is, we do not have cases in which predictable contexts are ended with unexpected completions. Next, for young adults, the effect is most evident for moderately (and not strongly) constraining contexts; the most predictable endings were not distinguishable from the unexpected items, a pattern that would be difficult to explain under a prediction account of the findings observed here. Finally, the scalp distribution of the two effects is not identical: the late positivity is largest over medial prefrontal sites, whereas the negativity is more left fronto-temporal.

variations in sentential context. In both groups, context has a facilitative impact on semantic processing as indexed by the N400. However, younger and older adults evidenced qualitatively different patterns of sensitivity over the entire range of contextual constraint. Whereas young adult N400s are very sensitive to availability of contextual information (measured by cloze probability and offline subjective ratings), older adults have a more categorical pattern, a change possibly resulting from a reduced reliance on an anticipatory mode of processing in comprehension. Future work should corroborate these claims with designs specifically meant to test for engagement of predictive comprehension mechanisms in the use of sentential contexts.

We also observed a later, left-lateralized frontal negativity (Wlotko & Federmeier, submitted; cf. Kutas, 1993; Federmeier & Kutas, 2005; Davenport & Coulson, 2011) for younger and older adults, which we linked to reinterpretation of the context, similar to frame-shifting. For the young, reinterpretation took place most notably in moderately constraining contexts, but for older adults, a similar negativity was observed also in strong contexts -- a change that may also be precipitated by a decreased reliance on a predictive mode of comprehension.

Taken together with past findings, we can conclude that there are multiple routes to comprehension, distributed across the hemispheres in the young adult brain and distinct across age. Indeed, the greater sensitivity to contextual information by young adults may arise out of the contributions from both hemispheres in parallel, whereas this “division of labor” may be less likely for older adults. Different processes are also engaged across time, with age-related changes in efficacy or availability of resources leading to shifts in reliance on the brain mechanisms recruited for comprehension. Reduced sensitivity to contextual information in earlier stages (N400) propagates to a greater reliance on later processes (the frontal negativity) for older adults. This type of pattern illustrates the brain's remarkable ability to flexibly adapt, through the use of distinct pathways, to diverse processing conditions and to achieve its goals using the resources available to it.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the McDonnell Foundation and NIH grant AG026308 to Kara Federmeier and NIH grant F31-DC009781 and an NIMH START-MH fellowship to Edward Wlotko. Scott Schulfer, Paul Ravenna, Andrew Dignan, and Cassandra Sung provided assistance with data collection and analysis for young participants. Charlotte Laguna, Caterina Gratton, and Jennifer Hanson collected the data and assisted with analysis from the older participants. Laura Giffin provided assistance with neuropsychological data for older participants. Tanya Stanley collected norming data from older participants. Yifei Du, Zoya Hussain, Sheela Gogula, Scott Schulfer, Sarah Baisley, Aubrey Lutz, and Neha Patel completed norming and neuropsychological data entry and scoring. Sarah Brown-Schmidt and Scott Fraundorf provided helpful comments on an earlier version of this manuscript.

References

Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974; 19(6):716–723.

- Benjamin AS. On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27(4):941–947.
- Brown-Schmidt S, Tanenhaus MK. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science: A Multidisciplinary Journal*. 2008; 32(4):643–684.
- Coulson S, Kutas M. Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*. 2001; 316(2):71–74. [PubMed: 11742718]
- Coulson S, Williams RF. Hemispheric asymmetries and joke comprehension. *Neuropsychologia*. 2005; 43(1):128–141. [PubMed: 15488912]
- Dale, AM. Unpublished Dissertation. San Diego, La Jolla, CA: University of California; 1994. Source localization and spatial discriminant analysis of event-related potentials: Linear approaches.
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of Memory and Language*. 1980; 19(4):450–466.
- Davenport T, Coulson S. Predictability and novelty in literal language comprehension: An ERP study. *Brain Research*. 2011; 1418:70–82. [PubMed: 21925647]
- DeLong KA, Urbach TP, Kutas M. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*. 2005; 8(8):1117–1121.
- Federmeier KD. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*. 2007; 44(4):491–505. [PubMed: 17521377]
- Federmeier KD, Kutas M. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*. 1999; 41(4):469–495.
- Federmeier KD, Kutas M. Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*. 2005; 42(2):133–141. [PubMed: 15787850]
- Federmeier KD, Kutas M, Schul R. Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*. 2010; 115(3):149–161. [PubMed: 20728207]
- Federmeier KD, McLennan DB, De Ochoa E, Kutas M. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*. 2002; 39(2):133–146. [PubMed: 12212662]
- Federmeier KD, Van Petten C, Schwartz TJ, Kutas M. Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging*. 2003; 18(4):858–872. [PubMed: 14692871]
- Federmeier KD, Wlotko EW, De Ochoa-Dewald E, Kutas M. Multiple effects of sentential constraint on word processing. *Brain Research*. 2007; 1146:75–84. [PubMed: 16901469]
- Forster KI. Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*. 1981; 33A(4):465–495.
- Goldstein, H. Multilevel statistical models. 4th ed.. West Sussex, UK: John Wiley & Sons, Ltd; 2011.
- Hagoort P, Van Berkum JJA. Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 2007; 362(1481):801–811.
- Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*. 2003; 49(1):133–156.
- King, JW.; Kutas, M. Do the waves begin to waver? ERP studies of language processing in the elderly. In: Allen, PA.; Bashore, TR., editors. *Age differences in word and language processing*. Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers; 1995. p. 314–344.
- Kluender R, Kutas M. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*. 1993; 5(2):196–214. [PubMed: 23972154]
- Kutas M. In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*. 1993; 8(4):533–572.
- Kutas, M.; DeLong, KA.; Smith, NJ. A look around at what lies ahead: Prediction and predictability in language processing. In: Bar, M., editor. *Predictions in the brain: Using our past to generate a future*. New York, NY: Oxford University Press; 2011. p. 190–207.

- Kutas M, Federmeier KD. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*. 2011; 62:621–647.
- Kutas M, Hillyard SA. Brain potentials during reading reflect word expectancy and semantic association. *Nature*. 1984; 307(5947):161–163. [PubMed: 6690995]
- Lahar CJ, Tun PA, Wingfield A. Sentence-final word completion norms for young, middle-aged, and older adults. *The Journals of Gerontology. Series B: Psychological Sciences and Social Sciences*. 2004; 59B(1):P7–P10.
- Laszlo S, Federmeier KD. A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*. 2009; 61(3): 326–338. [PubMed: 20161064]
- Laszlo S, Federmeier KD. The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*. 2011; 48(2): 176–186.
- Lee C, Federmeier KD. Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language*. 2009; 61(4): 538–555. [PubMed: 20161361]
- Light, LL.; Burke, DM. *Language, memory, and aging*. New York: Cambridge University Press; 1988.
- Meyer AM, Federmeier KD. Event-related potentials reveal the effects of aging on meaning selection and revision. *Psychophysiology*. 2010; 47(4):673–686. [PubMed: 20210876]
- Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 1971; 9(1):97–113. [PubMed: 5146491]
- Reuter-Lorenz PA, Jonides J, Smith EE, Hartley A, Miller A, Marshuetz C, Koeppel RA. Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of Cognitive Neuroscience*. 2000; 12(1):174–187. [PubMed: 10769314]
- Schwanenflugel PJ, LaCount KL. Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1988; 14(2):344–354.
- Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6(2):461–464.
- Simpson, GB., editor. *Understanding word and sentence*. Oxford, England: North-Holland; 1991.
- Snijders, TAB.; Bosker, RJ. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage Publications; 1999.
- Stanovich KE, West RF. Exposure to print and orthographic processing. *Reading Research Quarterly*. 1989; 24(4):402–433.
- Stine-Morrow EAL, Miller LMS, Hertzog C. Aging and self-regulated language processing. *Psychological Bulletin*. 2006; 132(4):582–606. [PubMed: 16822168]
- Swinney DA. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*. 1979; 18(6):645–659.
- Taylor WL. Cloze procedure: A new tool for measuring readability. *Journalism and Mass Communication Quarterly*. 1953; 30(4):415–433.
- Van Berkum JJA. Understanding sentences in context: What brain waves can tell us. *Current Directions in Psychological Science*. 2008; 17(6):376–380.
- Van Berkum JJA, Brown CM, Zwitterlood P, Kooijman V, Hagoort P. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(3):443–467.
- Van Berkum JJA, Hagoort P, Brown CM. Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*. 1999; 11(6):657–671. [PubMed: 10601747]
- Van Berkum JJA, Koornneef AW, Otten M, Nieuwland MS. Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*. 2007; 1146:158–171. [PubMed: 16916496]
- Van Petten C, Coulson S, Rubin S, Plante E, Parks M. Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1999; 25(2):394–417.

- Van Petten C, Kutas M. Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*. 1990; 18(4):380–393. [PubMed: 2381317]
- Van Petten, C.; Kutas, M. Electrophysiological evidence for the flexibility of lexical processing. In: Simpson, GB., editor. *Understanding word and sentence*. Oxford, England: North-Holland; 1991. p. 129-174.
- Wicha NYY, Moreno EM, Kutas M. Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*. 2004; 16(7):1272–1288. [PubMed: 15453979]
- Wilson M. MRC psycholinguistic database: Machine usable dictionary, version 2.00. *Behavior Research Methods, Instruments, and Computers*. 1988; 20(1):6–11.
- Wlotko EW, Federmeier KD. Finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*. 2007; 45(13):3001–3014. [PubMed: 17659309]
- Wlotko EW, Federmeier KD. So that's what you meant! A left-lateralized brain potential for sentences of moderately strong constraint. (submitted).
- Wlotko EW, Lee C, Federmeier KD. Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass*. 2010; 4(8):623–638. [PubMed: 20823949]
- Zacks R, Hasher L. Cognitive gerontology and attentional inhibition: A reply to Burke and McDowd. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*. 1997; 52B(6):P274–P283.

Appendix

Details of multi-level modeling

To examine N400 responses within and across age groups, linear mixed models with random intercepts and slopes (for subjects) were fit using PROC MIXED in SAS 9.2. For details about multi-level modeling, see Goldstein (2011) or Snijders and Bosker (1999). For each age group, the “null” model – with no predictor variable – was fit (i.e. a random effects ANOVA model where “groups” are subjects) to obtain estimates of between- and within-group variability (leading to the intra-class correlations). In both cases, the estimated between-subjects variance was large relative to its standard error, indicating the appropriateness of including random effects in the model. The appropriateness of including the random slope for cloze probability was tested explicitly by calculating the difference in deviance between the model with and without a random effect for cloze and comparing against a mixture of the two chi-squared distributions with degrees of freedom 1 and 2 (for both young and older, $p < .0001$). Table A1 shows parameters, standard errors, deviance, and information criteria (Akaike Information Criterion, Akaike, 1974; and Bayesian Information Criterion, Schwarz, 1978; i.e. “goodness of fit”) measures for the relevant models.

Parameters were estimated using restricted maximum likelihood estimation in order to obtain unbiased estimates of variance components and an “unstructured” covariance model. Fixed effects and confidence intervals were based on degrees of freedom estimated with the general Satterthwaite approximation and model-based estimates of standard errors. No conclusions were altered when compared with the results using empirical (rather than model-based) estimates of standard errors.

Separate models are presented for young and older groups, but a model with all participants and a fixed effect of group revealed no main effect of group (i.e. intercept; estimated

interaction parameter = $-.36$ for the young, $p=.5$, 95% confidence interval $[-.142, .71]$), but an interaction of group and cloze slope of 1.22 for young ($p<.0001$, 95% confidence interval $[.009, .015]$). This can be interpreted as revealing that overall, young adults have a steeper gradation of N400 amplitudes (across the full range of cloze, an effect of 2.18 for older adults and 3.4 for older adults, a difference of 1.22), consistent with the well-replicated result of smaller effects of sentential context on N400s for older adults. The estimated random effects of cloze with their 95% confidence limits for each participant as estimated from the combined model are presented in Figure A1. All young and all but one older participant have an estimated positive slope. In addition, most older participants have smaller slopes than most young participants.

Separate models for strong and weak constraint conditions were fit with Age Group as a factor. Parameters for these models are presented in Table A2.

Figure 1.

Grand average ERPs at all electrode sites for both age groups, collapsed into high (75–100%), medium (30–50%) and low (<30%) cloze probability bins. At center, the head icon (nose at top) indicates the approximate scalp position and the label of each electrode channel. Negative is plotted up in this and all subsequent figures.

Figure 2.

ERPs for all sentence-final words overlapped for young and older adults at left and right anterior, central, and posterior electrode sites.

Figure 3.

ERPs for the six 50-item cloze probability bins at 8 representative electrode sites. For younger and older adults, the N400 appears as a negative-going deflection in the waveform approximately 400 ms post-stimulus onset; a left-lateralized fronto-temporal negativity onsetting around 400–500 ms is observed for both groups, most enhanced in moderately strong contexts for young adults but also enhanced in very strong contexts for older adults.

Young Adults

Interpolated N400 item-level correlation values

Unexpected vs Moderately Strong Constraint
(600-900 ms)

Older Adults

Interpolated N400 item-level correlation values

Unexpected vs Moderately Strong Constraint
(600-900 ms)

Figure 4.

Scalp maps of interpolated correlation coefficients for N400 amplitudes and the voltage difference between Unexpected items and the 75–90% cloze bin in the 600–900 ms time window for younger and older adults. For both groups, the pattern of correlation over the scalp matches the well-known N400 distribution, and the later negativity for moderately constraining items produces a left fronto-temporal distribution.

N400 amplitude Older and Younger Adults

Figure 5.

Scatterplot of N400 averages for individual items versus cloze probability, plotted for both age groups. Young adult data appear in green/thin line and older adults appear in blue/thicker line in this all all following scatter plots. Older adults show less sensitivity in variation to contextual strength, particularly at weak levels of constraint.

Figure 6.

Scatterplot of item averages for the left frontal negativity versus number of alternative completions for the item's context, plotted for both age groups. Both groups show a larger negativity (relative to baseline) for items with fewer unique completions generated in a cloze norming task, but older adults evidence this relationship over a wider range of constraint.

Table 1

Examples of items from the conditions in the experiment presented with the best completion and examples of alternate completions from the cloze norming.

Cloze Bin	Sentence Frame	Best Completion (Cloze)	Alternate Completions (Cloze)
90 – 100	The little girl refused to go to sleep until he told her a	story (100%)	<i>n/a</i>
	Tricia had never seen a spider get tangled up in its own	web (97.1%)	legs (2.9%)
75 – 90 (few alternates)	He was cold most of the night and finally got up to get another	blanket (85.3%)	drink (8.8%), shirt (2.9%), sweater (2.9%)
	After two hours of hard work they decided to take a short	break (77.1%)	nap (17.1%), cut (2.9%), walk (2.9%)
75 – 90 (more alternates)	Don't touch the wet	paint (85.3%)	bed (2.9%), cement (2.9%), dog (2.9%), outlet (2.9%), sand (2.9%)
	I just had a new sound system installed in my	car (79.4%)	house (5.9%), garage (2.9%), home (2.9%), ride (2.9%), room (2.9%), truck (2.9%)
50 – 75	She pulled her head out from under the faucet and reached for a	towel (70.6%)	beer, cup, drink, napkin, pill, purse, screwdriver, toothbrush, wrench, <i>no response</i>
	Jim was saving boxes for a friend who was	moving (64.7%)	busy, recycling, away, collecting, dead, desperate, drunk, gone, poor, sick
10 – 50	He was caught stealing a	car (47.1%)	purse, wallet, base, baseball, butterfinger, coat, diamond, dog, erasers, fish, gun, horse, knife, lemon, rug, shirt
	It was time to hang the new	pictures (23.5%)	poster, banner, calendar, clock, clothes, coat, cord, decorations, drapes, frame, ornaments, painting, portrait, wreath, <i>etc.</i>
Unexpected	They waited a long time to see the	grades	
	Rushing out he forgot to take his	camera	

Table 2

Stimulus characteristics for each of the 50-item cloze probability bins.

	Primary Cloze	Alternate Cloze	Sentence Length	Kucera-Francis Frequency	Number of Letters	Concreteness	Familiarity	Imageability	Proportion of Nouns
90 – 100	94.2%	2.6%	10.0	118.4	4.7	520.7	576.1	552.1	0.84
75 – 90	83.2%	5.2%	10.2	110.8	4.7	519.6	578.5	541.1	0.80
50 – 75	64.6%	9.7%	9.7	115.0	5.1	541.9	577.3	559.8	0.86
30 – 50	38.8%	11.2%	9.6	110.0	4.8	527.6	579.4	542.7	0.82
10 – 30	21.7%	8.3%	9.7	118.5	5.2	520.0	569.3	543.5	0.80
Unexpected	1.4%	1.7%	10.1	113.1	5.4	492.6	571.3	524.3	0.82

Table 3

Expectancy and Plausibility ratings for both young and older adults.

	Expectancy - Young	Expectancy - Older	Plausibility - Young	Plausibility - Older
90 – 100	1.3	1.2	1.1	1.1
75 – 90	1.5	1.3	1.1	1.2
50 – 75	1.6	1.4	1.2	1.2
30 – 50	1.9	1.7	1.2	1.3
10 – 30	2.4	2.3	1.3	1.5
Unexpected	3.9	3.7	2.2	2.6

Table 4

Item-analysis regression results for younger and older adults.

cloze range	regression coefficient	confidence interval	F (df)	p value	R ²	adj. R ²
Young Adults						
10 – 90	cloze: 2.95	1.51, 4.38	F(1,198)=16.4	<.0001	0.08	0.07
10 – 50	cloze: 4.90	1.20, 8.60	F(1,98)=6.9	0.01	0.07	0.06
50 – 90	cloze: –0.5	–5.40, 4.41	F(1,98)=0.04	0.84	0	–0.01
Unexpected	cloze: 15.77	–1.72, 33.25	F(1,48)=3.29	0.07	0.06	0.04
	exp: –0.70	–1.29, –0.12	F(1,48)=5.78	0.02	0.11	0.09
	plaus: –0.76	–1.44, –0.08	F(1,48)=5.1	0.03	0.1	0.08
Older Adults						
10 – 90	cloze: 2.52	1.55, 3.48	F(1,198)=26.36	<.0001	0.12	0.11
10 – 50	cloze: 1.32	–1.31, 3.95	F(1,98)=0.99	0.32	0.01	0
50 – 90	cloze: –0.02	–3.12, 3.08	F(1,98)=0.0	0.99	0	–0.01