Age differences in language comprehension during driving: Recovery from prediction errors is more effortful for older adults

Katja Häuser (khaeuser@coli.uni-saarland.de)
Saarland University
Campus C7.4
66123 Saarbrücken
Germany

Vera Demberg (vera@coli.uni-saarland.de)
Saarland University
Campus C7.4
66123 Saarbrücken
Germany

Jutta Kray (j.kray@mx.uni-saarland.de)
Saarland University
Campus A1.3,
66123 Saarbrücken
Germany

Abstract

Prior research yielded conflicting findings regarding whether older adults show a greater processing cost than younger adults when encountering unpredicted semantic material during language processing. Here, we investigated whether age-related differences in recovery from prediction error are influenced by increased demands on working memory. We used a dual task design: a primary sentence comprehension task in which semantic predictions were fulfilled or violated, and a concurrent driving task, thought to limit working memory resources in resolving prediction errors. In the dual task, older participants showed an increase in comprehension accuracy for sentences with semantic violations, while demonstrating a decrease in driving accuracy. Thus, when working memory resources were limited, older adults focused exclusively on the language task and neglected the driving task. This could be related to an age-related increase in generating semantic predictions, or to a general inability among older adults to divide attention between two cognitively demanding tasks.

Keywords: aging, semantic expectancy, dual tasking, attention allocation

Introduction

Prediction of upcoming linguistic material is pervasive during language comprehension. Recent theories hold that expectations at higher levels of processing (e.g., syntactic, contextual) generate hypotheses and facilitate low-level processing, for example in word recognition (Kuperberg & Jaeger, 2016).

Frequently, however, people encounter unpredicted linguistic content and must recover from unexpected events that violate their expectations. Indeed, research has shown that this recovery phase often involves a processing cost. For example, Federmeier and Kutas (1999) analyzed the N400 EEG component (for review, see Kutas & Federmeier, 2000) to index comprehension difficulties when participants were reading unexpected sentence continuations. Sentences contained either an expected word, an unexpected word from the same semantic category, or an unexpected word from a different semantic category (e.g., *They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of* ***palms*** (expected) */* ***pines*** (unexpected same category) */* ***tulips*** (unexpected different category)). According to the results, the N400 was reduced for expected and semantically related words, indicating that processing of predicted (*palms*) and semantically related words (*pines*) was facilitated. In contrast, for semantically unexpected words (*tulips*) the N400 amplitude was high, suggesting comprehension difficulties among participants when predictions based on context were violated.

However, an unanswered question is whether older adults (65 years or older) use context to anticipate upcoming content during language processing in a similar fashion as younger adults do. Some studies have shown that older adults are more disturbed by unpredicted semantic material than younger adults, which suggests that older adults may rely more heavily on prediction making during language processing (DeLong, Groppe, Urbach, & Kutas, 2012; DeDe, 2014; Rayner, 2006; Borges & Coco, 2015). For example, Borges and Coco (2015) investigated age differences in visual object detection by using a priming paradigm in which prime and search scene were either congruent (e.g., *kitchen-kitchen*) or not (e.g., *bathroom-kitchen*). In addition, visual target objects (e.g., *bread basket*) were presented in a semantically consistent condition (e.g., on a restaurant table) or in an inconsistent condition (e.g., on a pool table). According to the results, older adults were less successful at detecting target objects when prime and search scene were semantically congruent, but the target was inconsistent with the search scene. The authors concluded that older adults rely more heavily on contextual expectations than younger adults by generating very specific predictions based on consistent information. Consequently, they showed a greater processing cost when expectations based on context and new information are inconsistent.

Other studies, in contrast, have shown that older adults generally appear less likely than younger adults to use context and engage in pre-activating information during sentence processing (e.g., Federmeier & Kutas, 2005; Federmeier, Kutas, & Schul, 2010). For example, Federmeier and Kutas (2005) compared younger and older adults' N400 amplitude for sentence-final words in highly and weakly constraining contexts (e.g., highly constraining: *No one at the reunion recognized Dan because he had grown a beard*; weakly constraining: *At the children's park next to the beach she saw a man with a beard*). Even though both age groups showed a similar N400 for weakly constraining sentences, the older adults' brain response for strongly constraining sentences was delayed and diminished in shape. The authors concluded that older adults were unable to make use of the richer information available from strongly constraining contexts to guide semantic processing; possibly because age-related declines in working memory prevented older adults from quickly constructing and updating an ongoing message-level representation while at the same time processing new input (see Huettig & Janse, 2016, for a similar account).

Given Federmeier and Kutas' (2005) implication of working-memory capacity, the goal of the present study was to examine whether age-related changes in predictive processing are influenced by increasing demands on working memory. We used a dual-task paradigm with a primary language comprehension task and secondary driving task, thought to limit cognitive resources that participants can expend to resolving semantic prediction errors. To our knowledge, only one previous study has investigated how aging affects dual-task performance during language processing and driving, and that particular study found an age effect that was limited in scope.

Becic *et al.* (2010) investigated story-retelling ability in younger and older adults while participants were engaged in a secondary driving task. According to the results, younger adults achieved high accuracy in both story retelling and driving, suggesting high capacity in this participant group to divide attention between the language and the driving task. For older adults, no reliable effects emerged in the primary analysis. However, there was a trend in the data (revealed by post-hoc tests), which suggested that better driving (less variability in velocity and lane keeping) was associated with worse retelling. In other words, older adults who drove better also performed more poorly in the language task. Since the group of older adults showed worse story-retelling ability overall, it seemed that older adults primarily focused on getting the driving task right, while neglecting the language task. The authors suggested that, due to age-related declines in working-memory capacity, older adults may be more likely to protect their driving by giving up on the story retelling task. However, the Becic *et al.* (2010) study remains somewhat mute with respect to age differences in predictive processing (the primary focus of the present study), since this question was not specifically addressed by that paper.

In the present study, we sought to adress age-related differences in recovery from prediction error more directly, by presenting stimuli in a low- vs. high-surprisal condition (e.g., *Since Petra didn't have anything to* ***wear*** *for the barbeque, she bought a* ***dress*** *(low surprisal) in a nearby shop*; vs. *Since Petra didn't have anything to* ***drink*** *for the barbeque, she bought a* ***dress*** *(high surprisal) in a nearby shop*). High-surprisal sentences were thought to induce a strong cognitive conflict since the second clause violated semantic predictions based on contextual information provided by the first clause (i.e., *drink-dress*).

Based on prior research on semantic surprisal in younger adults (DeLong, Troyer, & Kutas, 2014; Kutas & Federmeier, 2000), we predicted that younger adults should be sensitive to violations of semantic expectancies (probably indexed by lower accuracy for high-surprisal sentences in the language comprehension task). In addition, we expected stable performance in this participant group regardless of whether sentences were processed in the single or dual task, indicating high capacity in younger adults to divide attention even under conditions of high linguistic load (cf. Becic *et al.*, 2010).

In contrast, for older adults our predictions were less clear based on previous research. If, on the one hand, older adults generate more specific predictions during language processing, we expected to find large processing costs in response to high-surprisal sentences, in particular under dual-task conditions, when less cognitive resources are available to resolve the semantically unexpected event. If, on the other hand, older adults are less efficient at generating predictions, we expected only minimal processing costs for high-surprisal sentences, with only small differences between single and dual-task condition.

Method

Participants

Thirty-six older adults (*mean age* = 72 yrs; 18 female) from the Saarbrücken community participated for compensation. The control group consisted of 34 younger adults (*mean age* = 23; 20 female), mostly students at UdS. All participants were native speakers of German, reported no hearing problems and had normal or corrected-to-normal vision.

Procedure

We investigated age-related differences in recovery from prediction errors while participants were engaged in a single and dual task. The dual task consisted of a language comprehension and continuous driving task. The single task consisted of the driving or language task only. Overall, the experiment consisted of six major blocks – two dual-task blocks for simultaneous language comprehension and driving, two single-task blocks for single driving, and two single task blocks for language comprehension.

Language Task. The language comprehension task consisted of a sentence verification task for 192 spoken sentences, half of them presented in a low-surprisal condition (low processing effort) and high-surprisal condition (high processing effort). Each sentence was constructed of two clauses, with the verb of the first clause providing a semantic context and the noun of the second clause either matching (low-surprisal condition) or violating this semantic context (high-surprisal condition), for example, *Since Petra didn't have anything to* ***wear*** (low surprisal) ***/ drink*** (high surprisal) *for the barbecue, she bought a* ***dress*** *in a nearby shop*. Participants were instructed to carefully listen to the sentences, which were presented to them over speakers, and asked to judge whether each sentence was meaningful and correct by verbally answering "Yes" or "No", while the researcher recorded their responses. In order to minimize prosodic differences among items, all sentences were synthesized prior to the experiment using MARY TTS (Schröder, Charfuelan, Pammi, & Turk, 2008) and pauses manipulated so that the duration of the disambiguating word (*dress*) was always identical. To avoid stereotyped responses, we also presented 72 filler items in a low- and high-surprisal condition, involving syntactic violations. All items were randomized using a Latin Square randomization, with surprisal (high-low) as blocking factor, to ensure that each participant encountered each experimental item in only one of its experimental conditions.

Driving Task. We used the Continuous Tracking and Reaction Task (ConTRe Task; Mahr, Feld, Moniri & Math, 2012), a highly controlled driving task which measures rapid changes in steering deviation from a target. As such, the ConTRe task allows for continuous and very fine-grained measurement of online changes in task performance over time (e.g., Becic, Dell, Bock, *et al.*, 2010; Demberg, Sayeed, Mahr, & Müller, 2013). Participants were seated in front of a steering wheel and saw a 3D road on a screen, with two vertical color bars moving laterally across the screen at a continuous speed. Participants were instructed that they could control one of the bars (the blue one) by turning the steering wheel whereas the other bar (the yellow one) was controlled by the computer. Their task was to continuously track the yellow bar so as to keep the distance between the two bars minimal at all times. Participants' driving performance was assessed by measuring their steering deviation (indicated in meters) when processing low- and high-surprisal sentences.

Results

We constructed separate linear mixed effects models for response accuracy and steering deviation, as implemented in the lme4 library (Bates & Sarkar, 2007) in R (R Development Core Team, 2013). Fixed effects for response accuracy were sentence type (low surprisal vs. high surprisal), task condition (single vs. dual), and age group (younger vs. older). Fixed effects for steering deviation were sentence type and age group. Since raw steering deviation was coded in positive and negative values, indicating left- and right-sided deviations, we squared its values to obtain a final measure. For the LMER model for steering deviation, *p*-values were approximated from the model coefficients using the normal distribution (see Barr, Levy, Scheepers, & Tily 2013). Categorical predictors were sum coded. All models contained participants and items as crossed random effects, and random slope adjustments for subjects and items. In the event that a model failed to converge, we simplified the random slope structure progressively until convergence was achieved (for guidelines, see Barr *et al.*, 2013). Higher-order interactions involving the factor age group were followed up with planned model splits between younger and older adults.

Response accuracy

The model for response accuracy showed a significant interaction between sentence type and age group, as well as a significant interaction between sentence type and task condition (see Table 2). To locate the source of these interactions, we computed two follow-up models in which items were split by age group. Thus, we computed one model for younger adults, and another model for older adults.

The model for the younger adults showed nothing but a significant main effect of sentence type (b = -0.62, SE = 0.14, t = -4.29, p < .001 ***), indicating that, regardless of task condition, younger adults responded less accurately to high-surprisal than low-surprisal sentences (see Figure 1, left panel). In contrast, the model for the older adults showed a significant interaction between sentence type and task condition (b = -0.86, SE = 025, t = -3.4, p < .001 ***). An inspection of the plot for this interaction (see Figure 1, right panel) suggested that older adults responded equally accurately to high- than low-surprisal sentences in the single task condition, but showed a selective increase in response accuracy for high-surprisal sentences in the dual-task condition. These observations were confirmed by additional follow-up models, in which we split items by task condition: As predicted, only the model for the dual-task condition showed a significant main effect of sentence type (b = 0.77, SE = 0.18, t = 4.28, p < .001 ***), indicating an increase in response accuracy for high-surprisal sentences.

Thus, the data for response accuracy showed two main things of interest: First, younger adults responded less accurately to high-surprisal sentences, regardless of task condition, indicating stable performance in this participant group even when working-memory load was high (i.e. the dual-task condition). Older adults, in contrast, responded more accurately to high-surprisal sentences in the dual-task condition, indicating that they selectively focused on resolving the semantic conflict in these items (cf. Becic *et al.*, 2010), presumably by giving up driving. To support this view, we now turn to the driving performance in both age groups.

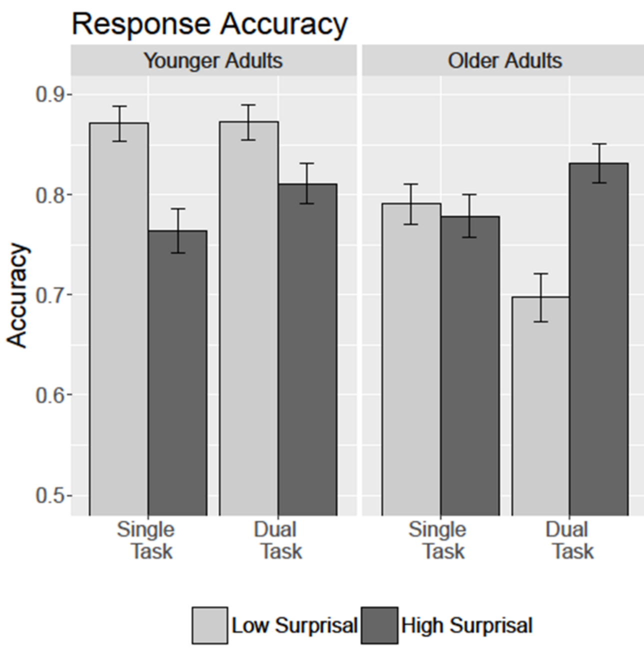

Figure 1: Response accuracy (± SEM) in younger and older adults, depending on task condition and sentence type.

Table 1: Effect sizes (b), standard errors (SE), t-values, and p-values for the logistic LMER model for response accuracy. Significance codes: *** .001 | ** .01 | * .05

	b	*SE*	*t*	*p*
Sentence Type	-0.12	0.10	-1.26	*ns*
Task Condition	-0.03	0.10	-0.35	*ns*
Age Group	0.34	0.15	2.24	*
SentType:Task	-0.58	0.19	-3.03	**
SentType:Group	-0.98	0.19	-5.11	***
Task:Group	-0.24	0.19	-1.27	*ns*
SentType:Task:Group	0.59	0.38	1.53	*ns*
Random Effects		*Variance*		
Subject		0.23		

Steering deviation

The model for squared steering deviations showed a significant interaction between sentence type and age group (see Table 2). The plot of this interaction (see Figure 2) suggested that younger adults showed stable driving performance regardless of whether sentences were highly surprising or not, whereas older adults demonstrated higher steering deviations when high-surprisal sentences were presented. To confirm these observations, we again computed follow-up models in which we split items by age group. As expected, only the model for older adults showed a main effect of sentence type (b = 0.24, SE = 0.06, t = 4.05, p < .001***; younger adults: b = -0.02, SE = 0.03, t = -0.70, p > .05).

Thus, the analysis of the driving data supported our hypothesis based on the response data. First, younger adults showed constant steering deviations regardless of semantic violations, suggesting that even under conditions of high linguistic load, they maintained high driving acuity. Older adults, in contrast, demonstrated greater steering deviations in response to high-surprisal sentences, suggesting increased effort to recover from semantic violations.

In sum, whereas younger adults maintained a stable pattern of performance even under conditions of high linguistic load, older adults devoted all attentional resources to resolving semantic violations, while neglecting the driving task.

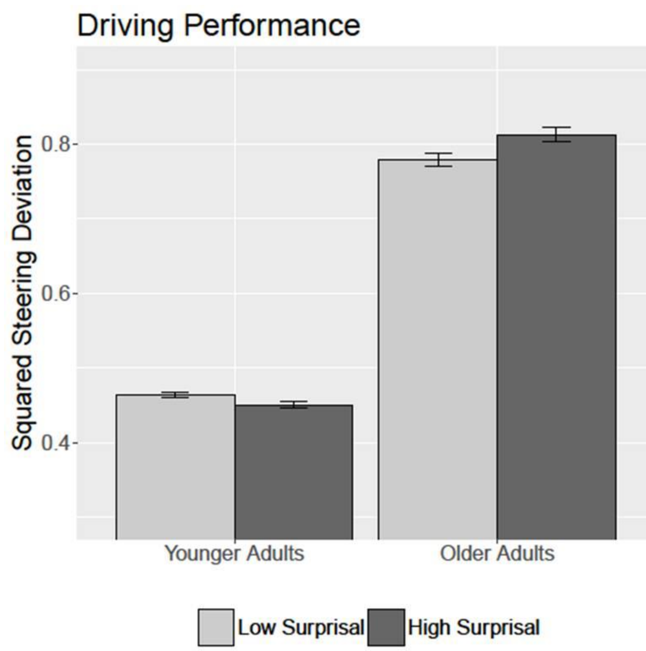

Figure 2: Steering deviations in the dual task (± SEM) for younger and older adults, depending on sentence type.

Table 2: Effect sizes (b), standard errors (SE), t-values, and bootstrapped p-values for the logistic LMER model for steering deviation.

	b	*SE*	*t*	*p*
Sentence Type	0.11	0.03	3.53	***
Age Group	0.31	0.17	1.80	*ns*

SentType:Group	0.26	0.06	4.18	***
Random Effects		*Variance*		
Subject		0.50		
SentType | Subject		0.06		
Item		0.01		
Age Group | Item		0.05		

Discussion

Prior research has yielded conflicting findings with respect to predictive processing in aging. Some studies have shown that older adults are impaired at using context to generate predictions about upcoming content during language comprehension. Other studies have indicated that older adults form strong and semantically specific predictions during language processing, resulting in effortful recovery when such predictions are violated.

In this study, we investigated age-related differences in recovery from prediction error under conditions of increased working-memory load. By using a secondary driving task, we limited working-memory resources participants could devote to resolving prediction errors. To manipulate prediction error, we presented sentences in a high- and low-surprisal condition. In high-surprisal sentences, participants were expected to experience integration difficulties when encountering unpredicted semantic content. Low-surprisal sentences, in contrast, were thought to induce only minimal processing effort.

Two key findings emerged. First, even though younger adults were sensitive to violations of semantic predictions overall (indicated by lower response accuracy for high-surprisal sentences), they maintained a stable behavioral pattern in both response accuracy and driving performance. Thus, younger adults were able to resolve the semantic violation in high-surprisal sentences without experiencing trade-off effects between primary and secondary task. This suggests high working-memory capacity in this participant group to split attention even under conditions of maximal linguistic load. Second, we found that older adults allocated all processing resources towards resolving the unexpected sentence continuation in high-surprisal sentences. This increased their response accuracy in the sentence verification task, but it came at the expense of driving accuracy: When high-surprisal sentences were presented, older adults demonstrated a strong increase in steering deviation.

Thus, our results are more in line with studies suggesting that older adults form strong predictions during language processing, and that violations of these predictions induce maximal processing effort to resolve the prediction error. Unlike younger adults, however, older adults may not have sufficient working-memory capacity to integrate semantically unexpected material into an unfolding sentence context and to additionally perform a secondary task without a substantial drop in task performance.

A second interpretation of our results is that older adults are inable to successfully divide attention between two cognitively demanding tasks. Thus, they might globally shift attentional resources towards one cognitive goal when multiple tasks have to be performed at the same time. This interpretation is in line with prior research suggesting that older adults can relevel their task priorities in a case-by-case manner that follows principles of selective optimization, by taking into account the subjective difficulty of each task and choosing the one which is most likely to garner success (see Li, Baltes, Staudinger, & Lindenberger, 1999; Miles & Stine-Morrow, 2004; Stine-Morrow, Miller & Hertzog, 2006). Here, older adults might have adopted a strategy of selective performance optimization, by neglecting the high demands in the bar-tracking task and focusing exclusively on the sentence verification task. Overall, the language task may have seemed more likely to yield success, given older adults' increased verbal knowledge and linguistic capacity (Glisky, 2007).

Finally, a somewhat open question is to what extent our data have real-life implications on older adults' car driving security. On the one hand, our results are supported by prior studies using simulated but also naturalistic driving scenarios, suggesting that driving ability suffers most under conditions of high working-memory load (Cantin, Lavallière, Simoneau, *et al.*, 2009; Strayer, Cooper, Turrill, *et al.*, 2013), and that older adults are more likely to adapt to such situations by selectively focusing their attention on one task and disregarding the other (Becic *et al.*, 2010). In addition, there is evidence from research on car driving safety (Strayer *et al.*, 2013) indicating that behavioral results obtained in simulated driving environments are largely identical to real-life driving.

On the other hand, car driving involves a range of cognitive-behavioral demands the bar tracking task in the present study was lacking, for example reactions to road signs or traffic lights, overtaking maneuvers and lane changes, or braking for other cars. In fact, the contrast between the present results and those obtained by an earlier study using a similar dual-task paradigm (Becic *et al.*, 2010), shows that differences in design might impact the results to a large extent. According to the results of Becic *et al.* (2010), older adults showed a reversed pattern of task prioritizing than observed here, by focusing exclusively on the driving task and neglecting the language task. However, the experimental set-up in Becic *et al.* (2010) was more naturalistic than the one in the present study, since that study used an actual car dummy and a wrap-around projection screen that displayed realistic images of road situations and naturalistic driving scenarios. This set-up might have induced a more realistic feeling of car driving, where accidents can actually be fatal. The older adults in that study might have employed a task solving strategy that followed the rule of safety-first, by focusing their attention on the task which seemed most dangerous to them. In contrast, poor performance in the bar tracking task in the present study had no such real-life implications, possibly rendering this task somewhat negligible to older adults.

In sum, our data support studies arguing that recovery from prediction errors is more effortful for older adults, and that older adults allocate attentional resources differently from younger adults when task demands are high, by prioritizing one cognitive goal over others. We discussed two possible causes for these age-related differences, i.e. older adults' increased rate of forming semantic predictions based on context, and/or impaired working-memory resources normally associated with aging. Future work in our lab will further investigate these possibilities by also exploring the pupillary response as a measure of cognitive load, and by taking into account individual differences in executive functions.

References

Bates, D.M. & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes, R package version 0.99875- 6.

Becic, E., Dell, G. S., Bock, K., Garnsey, S. M., Kubose, T., & Kramer, A. F. (2010). Driving impairs talking. *Psychonomic Bulletin & Review*, *17*, 15-21.

Borges, M., & Coco, M. I. (2015). Access and Use of Contextual Expectations in Visual Search during Aging. In G. Airenti, B. Bara, & G. Sandini (Eds.). *Proceedings of the EuroAsianPacic Joint Conference in Cognitive Science.*

Cantin, V., Lavallière, M., Simoneau, M., & Teasdale, N. (2009). Mental workload when driving in a simulator: Effects of age and driving complexity. *Accident Analysis and Prevention*, *41*, 763-771.

DeDe, G. (2014). Sentence comprehension in older adults: evidence for risky processing strategies. *Experimental Aging Research*, *40*, 436-454.

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*, 226-239.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre- processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*, 631-645.

Demberg, V., Sayeed, A., Mahr, A., & Müller, C. (2013). Measuring linguistically-induced cognitive load during driving using the ConTRe task. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 176-183). ACM.

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

Federmeier, K. D., & Kutas, M. (2005). *Aging in context: Age-related changes in context use during language comprehension*. *Psychophysiology*, *42*, 133–141.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*, 149-161.

Glisky, E. L. (2007). Changes in cognitive function in human aging. In D. R. Riddle (Ed.), *Brain Aging: Models, Methods, and Mechanisms*. New York: CRC Press.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition & Neuroscience*, *31*, 80-93.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32-59.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future*. New York: Oxford University Press.

Li, K. Z., Lindenberger, U., Freund, A. M., & Baltes, P. B. (2001). Walking while memorizing: Age-related differences in compensatory behavior. *Psychological Science*, *12*, 230-237.

Mahr, A., Feld, M., Moniri, M. M., & Math, R. (2012). The contre (continuous tracking and reaction) task: A flexible approach for assessing driver cognitive workload with high sensitivity. *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 88–91). ACM.

Miles, J. R., & Stine-Morrow, E. A. (2004). Adult age differences in self-regulated learning from reading sentences. *Psychology and Aging*, *19*, 626.

R Development Core Team. (2013). R: A language and environment for statistical computing. [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*, 448-465.

Schröder, M., Charfuelan, M., Pammi, S., & Türk, O. (2008). The MARY TTS entry in the Blizzard Challenge 2008. In Proceedings of the Blizzard Challenge

Stine-Morrow, E. A., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin*, *132*, 582.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile.