Neural Adaptation to Speaker Reliability and Disfluencies in Speech:

An Individual Differences Study

A Senior Honors Thesis for the Biopsychology Major, Department of Psychology

Rebecca Nardulli

Tufts University

Advisor: Gina Kuperberg, M.D., Ph.D.

Thesis Committee: Gina Kuperberg, M.D., Ph.D., Ariel Goldberg, Ph.D., Meredith Brown, Ph.D., Nathaniel Delaney-Busch, Ph.D.

Abstract

Past research has shown that cognitive differences between individuals impact the way we process and comprehend language (Shah & Miyake, 1996; Boudewyn, 2015). Specifically, individuals with greater executive function, as indexed by better performance on cognitive tasks such as the AX-Continuous Performance Task, have been found to have stronger receptive vocabularies, grammatical abilities, and natural language comprehension (Daltrozzo, Emerson, Deocampo, Singh, Freggens, Branum-Martin, & Conway, 2017; Misyak & Christiansen, 2011). These differences impact the way that we, as listeners, make predictions about upcoming lexical items. Disfluency also influences our predictive abilities (Arnold, Hudson Kam, & Tanenhaus, 2007). The present ERP study of the N400 effect following disfluency investigates how people's predictions about sentence continuations are mediated by the presence of disfluency, cloze probability, and implicit knowledge of the speaker's reliability (Brown et al., 2017). This exploratory analysis aims to use neuropsychological measures indexing constructs such as cognitive inhibition, statistical learning, personality type, and working memory to explore the sources of individual differences in both prediction and adaptation as found in this ERP study. While the results of our analysis indicated that any variation in prediction and adaptation found across participants could not be explained by differences in the neuropsychological constructs tested here, it is clear to us that there still remains much more investigation to be done.

Introduction

DISFLUENCIES, PREDICTION, AND LEARNING

Prediction is a central component of language comprehension. Efficient language processing comes from being able to listen to what has been said and make a probabilistic prediction of what, from this context, might come next (Federmeier, 2007). When listeners are faced with an input that they have not predicted, the resulting prediction error may lead them to learn and adapt, refining their predictions so that they might be more successful in the future (Kuperberg & Jaeger, 2015).

Disfluency, or irregularity in the flow of fluent speech, is one factor that might impact a listener's lexical predictions. Previous research has found that speakers utilize disfluencies systematically. Most commonly, they are used when speakers are trying to summon a word that may be difficult or unexpected, or need an extra moment to figure out what they are trying to say (Beattie & Butterworth, 1979; Clark & Fox Tree, 2002). Due to this systematicity of production, it can be inferred that disfluencies might be used by listeners systematically as well, leading to predictions about what the speaker might say next. For instance, if a speaker were to say, "Please pass the salt and *uhh…*" a listener would be unlikely to predict the continuation *pepper* due to its high predictability. However, if after making this prediction the listener finds that she was incorrect, and the speaker *did* mean to say "pepper," then she might be inclined to adjust her expectations about the speaker's use of disfluency in the future, thus abandoning her previous assumptions.

Previous research has shown that listeners also adjust their expectations, and thus predict differently, when they are given explicit information about a speaker that implies that they might use disfluency irregularly. For instance, Arnold, Kam, and Tanenhaus (2007) found in an eye tracking experiment that when told that the speaker had object agnosia, listeners were less likely to use the speaker's disfluencies as cues for which object he or she was attempting to describe than when they were not given any information about the speaker at all. This exhibited ability to adapt is the basis of the present ERP study, where we investigated how listeners' use of disfluencies (in this case, the word "uh") might adapt to implicit, or learned, information about the reliability with which a speaker produces them.

We look to the N400 ERP component to study this response. The N400 is a negative-going event-related brain potential response linked to meaning processing (Kutas & Hillyard, 1980). Its amplitude is modulated by the predictability of the preceding context: if an incoming word is highly supported by its context, then the N400 amplitude is attenuated (Kutas & Federmeier, 2011). However, if an incoming word is *not* supported by its context, then the amplitude will be larger. For instance, in our previous example, "Please pass the salt and..." the continuation "pepper" (the expected completion of this sentence) would evoke an attenuated N400 amplitude compared to a continuation such as "ketchup," as the context has already created a very strong prediction for "pepper." Listeners have little difficulty semantically processing "pepper," as it does not violate their previous assumptions. Thus, it is easily integrated into the preexisting context. The N400 effect (defined as the difference in amplitude between an expected and unexpected completion) has been well-established as representing the extent to which comprehenders' semantic expectations have been met, making it an asset in studying graded prediction at the semantic and lexical levels (Kuperberg & Jaeger, 2015).

Previous work has also been done to link the N400 effect to the comprehension of disfluencies. For instance, Corley et al. (2008) conducted a study requiring participants to listen to sentences that ended in either predictable or unpredictable words. They found that the N400 difference between predictable and unpredictable words was attenuated when words were preceded by a disfluent silence, giving rise to the “expect the unexpected” hypothesis that suggests that the presence of disfluency might cause listeners to have less faith in their lexical predictions (or abandon them entirely), resulting in a lesser prediction violation (and thus a lesser N400 amplitude). The results found by Arnold et al. (2007) also support this hypothesis. However, a second hypothesis surrounding the use of disfluencies has also arisen, suggesting that the disfluency is merely used as an attention-orienting cue that actually signals listeners to *increase* the strength of their predictions about upcoming content (Fraundorf & Watson, 2011). The results found in our lab support this second hypothesis.

We chose to use an individual differences approach to analyze the N400 effect found in the current study of disfluency. EEG research has broadened our knowledge of language processing immensely, far beyond what would be possible using behavioral studies alone. This work has helped to illuminate the great amount of variability in how the brain processes language. With the understanding that language processing engages many general cognitive processes, such as attention, working memory, and cognitive control processes, researchers have begun to focus on how individuals differ in these respects, and how those differences impact the ability to process and comprehend language (Boudewyn, 2015).

Here, we aim to use neuropsychological measures indexing these various constructs to explore the sources of individual differences in both prediction and adaptation as found in the described ERP study of the effects of disfluencies and speaker reliability on listener expectations. We then analyzed the results using data from neuropsychological testing conducted on the same participants in a second experimental session.

ERP EXPERIMENT: SUMMARY

The present study contained three manipulations, as detailed below:

Word Expectancy and Disfluency. The first manipulation addresses the relationship between expectancy and fluency. To do this, Brown et al. created three "types" of stimuli, each with and without the addition of disfluency:

(1a) Highly constraining context, highly expected ending
(1b) Highly constraining context, unexpected ending
(1c) Low constraint context, unexpected ending
(2a) Highly constraining context, disfluency, highly expected ending
(2b) Highly constraining context, disfluency, unexpected ending
(2c) Low constraint context, disfluency, unexpected ending

This analysis will focus on the effects of expectancy with high sentential constraint (i.e. items 1a, 1b, 2a, and 2b). In Storch (2016), it was hypothesized that disfluencies followed by a low-cloze word would have an attenuated N400 amplitude compared to disfluencies followed by a high-cloze word, and that the difference between the two conditions would be smaller when disfluencies were present.

Speaker Reliability. The second manipulation addresses the listener's implicit knowledge of the speaker. With the knowledge from previous studies that listeners use disfluencies differently when they have explicit knowledge of the speaker's reliability, Brown et al. manipulated the proportion of trials in which disfluency preceded unpredictable versus predictable words. To do this, participants were split into two groups: one that heard a reliable speaker who always used disfluencies in the expected manner (preceding an unexpected ending), and another that heard a speaker who was equally as likely to use disfluencies in front of an expected word as she was to use them in front of an unexpected one. It was hypothesized that participants who heard the unreliable speaker would gradually learn that this speaker was unreliable, and begin to disregard the disfluency as a clue for prediction modulation.

ERP EXPERIMENT: RESULTS

Contrary to what was originally hypothesized and found in the preliminary analysis done by Storch (2016), we found that the N400 amplitude was actually larger for words in disfluent contexts than it was for words in fluent contexts. These results are more consistent with the hypothesis suggested in studies done by researchers such as Fraundorf and Watson (2011), who surmise that listeners use disfluencies as attention-orienting cues rather than using them to inform or modulate the semantic content of their predictions. While we cannot determine whether the larger N400 effect that arises from disfluency in our study is due to comprehenders strengthening their predictions about upcoming content, we can confirm that listeners do use disfluencies systematically.

THE INDIVIDUAL DIFFERENCES APPROACH

Past research has found ample evidence that differences between individuals affect language processing and comprehension. For instance, we know that processing and storage components of working memory tasks are good predictors of performance on language processing tasks (Shah & Miyake, 1996); we also know that implicit statistical learning predicts comprehension accuracy (Misyak & Christiansen, 2011), and natural language ability (Daltrozzo et al., 2017).

In recent years, a growing number of studies using EEG have also been conducted to study the neural basis of language processing and variability between individuals. For instance, studies have related variability in performance on tasks involving cognitive processes such as word-decoding (Perfetti, Wlotko, & Hart, 2005) or working memory (Nakano, Saron, & Swaab, 2010) to ERP components such as the N400, with findings suggesting that electrophysiological studies of variability between individuals only continue to illuminate the ways in which neuropsychological differences influence language processing (Boudewyn, 2015).

Here, we seek to use the individual differences approach to explain and analyze a subset of results from the ERP study described above and in Brown, Delaney-Busch, Storch, Wlotko, & Kuperberg (2017). By doing so, we hope to attribute some of the differences in prediction and adaptation found between individuals to the differences in cognitive ability and executive function that exist between them.

The neuropsychological tests and surveys included in this study were intended to measure four different constructs: cognitive inhibition, statistical learning, personality type, and working memory. In addition, participants supplied information on their language learning which will be used in a brief discussion of the possible effects of bilingualism later in the discussion.

It is notable that the modality and domains of the tasks used to evaluate participants sometimes varied from the ones used in the ERP study of disfluency; for instance, while the study stimuli were presented aurally, many of the tasks were visual. It cannot be said with certainty that a task engaging the visual system, such as the Visual Statistical Learning task, will predict use of auditory information. This concern will be addressed later in the discussion section of this paper.

NEUROPSYCHOLOGICAL MEASURES

Cognitive Inhibition

Inhibitory resources are thought to be recruited in such a way that unwanted information can be prevented from coming to mind (Anderson, 2003). According to work done by Gernsbacher (1990, 1997) and Gernsbacher & Faust (1991), the ability to suppress context-irrelevant information is a necessary precursor to being able to efficiently construct message-level representations (Boudewyn, 2012). Cognitive inhibition tasks such as the AX-Continuous Performance Task and the Stroop Test measure an individual's ability to suppress task-irrelevant information such as predictions (as in the AX-Continuous Performance Task) or automatic processing (as in the Stroop Test). This is thought to influence how individuals respond to prediction violations.

AX-Continuous Performance Task. The AX-Continuous Performance Task (AXCPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956) is a test of cognitive control in which participants are given a target sequence (AX) and then presented with a continuous string of letter pairs. They must respond using one key ("1") for all trials in which they see an A followed by an X, and with another key ("2") for all other trials, which include the letter pairs AY, BY, and BX. One notable difference between the AXCPT task and the task that participants are asked to do in the present ERP study is that AXCPT occasionally requires participants to suppress their predictions after the stimulus is presented (i.e. know that they must press "2" when presented with an A followed by a Y, despite anticipating the target), whereas the present ERP study might require participants to suppress their predictions *before* the stimulus is presented once they have learned of the speaker's reliability. However, the same suppression mechanisms would still be employed.

AXCPT consists of several different trial types (AX, AY, BX, and BY) which can be compared to each other in order to measure several different constructs. AX is the target; participants must maintain a goal in order to continue to respond to AX trials. AY trials measure cognitive control by requiring participants to suppress the target expectancy and correctly identify the sequence as AY. How much that conflict slows down their reaction time can be seen as a measure of participants' abilities to inhibit their predictions for the upcoming stimulus (later referred to as AX:AY differentiation). BX trials, on the other hand, measure participants' abilities to activate and maintain the B-cue, allowing them to correctly identify BX rather than seeing _X and assuming the target. Finally, BY trials are used as an internal baseline measure of general performance ability (Braver, 2013).

For this analysis, we will focus on two measures of proactive control, or the ability to maintain a goal and use it to make predictions (for instance, expecting an X when given an A due to knowledge of the target). The first measure we use is the difference in reaction time between AY and AX trials. This allows us to isolate proactive slowdown and use it to assess predictive processing: if participants have a significantly longer reaction time in response to AY trials than they do AX trials, then there is evidence that they are exercising a proactive or predictive strategy to the task. In a similar vein, we calculated the Proactive Behavioral Index (PBI), or the degree to which participants showed evidence of proactive strategies being relied upon more than reactive strategies ("reactive" meaning seeing an X and thinking back to whether the last letter presented was an A or not in order to make a decision). For both cases, we hypothesize that participants who show greater proactive control in the form of either greater latency to AX:AY differentiation or a greater PBI will tend to make stronger predictions during language comprehension, and thus will show an increased N400 expectedness effect due to a greater impact of prediction violation.

Stroop Test. The Stroop test (Stroop, 1935) requires participants to respond to stimuli words based on the color in which they are written while ignoring what the words actually say (e.g. pressing the keyboard key corresponding to “blue” when presented with the word “red” in blue typeface). Participants with faster reaction times to incongruent trials compared to congruent trials during this task are better able to suppress task-irrelevant information (in this case, what the word actually says). They are less slowed down by incongruence between what they see (the word) and what they aim to identify (the color), and thus are better able to override conflict. This implies that they will show an attenuated N400 expectedness effect regardless of the item’s fluency, since they will be less impacted by prediction violation than participants with a greater difference between reaction times on congruent and incongruent trials (later referred to as the Stroop slowdown) during the Stroop task.

Additionally, it is hypothesized that individuals with less Stroop slowdown will show a lesser effect of disfluency on the N400 expectedness effect. Less Stroop slowdown means that participants are better able to suppress task-irrelevant information. Therefore, when presented with a sentence containing a disfluency followed by an expected continuation, participants will be better able to suppress the attention orientation that occurs as a result of the disfluency, thus attenuating the influence that disfluency has on the N400 expectedness effect.

Previous work done by Boudewyn, Long, and Swaab (2012) has shown that individuals who perform poorly on the Stroop task, and thus have poorer suppressive abilities, show larger N400 effects of lexical association, indicating that an individual’s general suppression ability contributes to individual differences in sensitivity to word-level associations during sentence comprehension. These findings, as well as the idea that the informative disfluencies should be prompting people to partially suppress the predicted continuation, lend credence to the idea that the attenuated N400 effect that results from speakers' use of disfluencies may be attributed to or impacted by individual differences in cognitive inhibition.

Statistical Learning

Statistical learning is thought to be a means by which individuals learn regularities within linguistic input, thus enabling language acquisition (Daltrozzo et al., 2017). The group effect of the present ERP study is designed to see if listeners will adapt to the statistics of the disfluency usage, or learn the speaker's "regularities." It is proposed that this domain-general visual statistical learning measure could relate to the rate (or size) of this adaptation.

Visual Statistical Learning. The Visual Statistical Learning task (VSL; Fiser & Aslin, 2002) requires participants to view a series of 288 images that are presented on the computer screen. There are twelve unique images in this series, and they are presented in groups of three, although participants are not aware of this grouping. After this exposure phase, participants are presented with two groups of triplets at a time (one they have seen before and a foil triplet that was never presented) and must decide using two keyboard keys which triplet is more familiar to them. Each triplet is tested eight times, for a total of 32 test trials within this phase. In order to be successful at this task, participants must implicitly learn about these groups of images and then use this learned information to identify the triplets that they have seen before. It is hypothesized that participants with higher scores on the VSL task will show greater adaptation effects over the course of the experiment, since they will be more able to use implicitly learned information about the speaker to determine how they should treat that speaker's disfluency. Once they learn to disregard the unreliable speaker's disfluency as a cue for prediction, they will have weaker predictions overall, leading to an attenuated N400 expectancy effect.

Personality Type

Autism Spectrum Quotient. The Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) is a self-report scale that measures the degree to which an adult with normal intelligence has the traits associated with the autism spectrum. Compared with the general population, individuals with high-functioning autism (IQ>70) have been found to be significantly better at processing single words than processing the meanings of complex sentences (Goldstein, Minshew, & Siegel, 1994). In addition, autistic individuals tend to show less pragmatic competence than age-matched controls (Baron-Cohen, 1988). Because disfluency is a social and pragmatic cue, it is likely that individuals with high autism quotients might not pick up on its implications. Further, the results of an fMRI study by Just (2004) investigating cortical activation and synchronization during sentence comprehension in high-functioning autism suggest that autistic individuals engage less in the integrative aspects of sentence processing than do individuals without autistic traits, lending credence to the idea that individuals who exhibit more traits associated with the autism spectrum might show less adaptation to the speaker's disfluencies over the course of the experiment than individuals with fewer traits associated with the autism spectrum.

Schizotypal Personality Questionnaire. The Schizotypal Personality Questionnaire (SPQ; Raine, 1991) is a self-report scale for the assessment of schizotypal personality disorder based on DSM-III-R criteria. Individuals with schizophrenia have been found to have numerous linguistic impairments with respect to both comprehension and production. Compared to speech produced by individuals without schizophrenic traits, the predictability of speech produced by schizophrenic patients is relatively low, as is their ability to make predictability judgments on normal speech (Kuperberg & Caplan, 2003). Acutely psychotic patients have also been found to perform poorly in tasks evaluating the acceptability of semantically anomalous sentences, suggesting that individuals who exhibit more schizophrenic traits might be less able to detect the relative “strangeness” of unexpected continuations (Anand, Wales, Jackson, & Copolov, 1994). Additionally, schizophrenic patients have been found to exhibit poorer proactive control than do healthy controls (Lesh et al., 2013), making them less likely to plan on using context in order to generate a strong lexical prediction. Thus, we hypothesize that participants with greater SPQ scores will show an attenuated N400 expectedness effect.

Working Memory

General working memory is thought to be particularly important for complex cognition, including language processing. Differences in the amount of information individuals can keep in working memory have been found to predict language acquisition as well as language processing (Tsai, Au, & Jaeggi, 2016). Importantly, verbal working memory has been found to predict verbal ability in young adults (Shah & Miyake, 1996). Additionally, verbal working memory ability has been shown to be a good predictor of older adults’ ability to use contextual information for the recognition of upcoming words while listening to sentences (Janse & Jesse, 2014). Although a recent study done by Boudewyn (2012) examining the effects of individual differences in working memory on sensitivity to lexical association found that working memory was not a sufficient predictor in this case, we know that verbal working memory has been shown to predict neural measures of language processing in many other instances (Boudewyn, Long, &

Swaab, 2013; Nakano et al., 2010; Van Petten, 1997). As such, we used two different measures of working memory to assess individuals' general ability to keep information online: one lexical and one numerical.

Listening SPAN. The Listening SPAN task (Daneman & Carpenter, 1980) assesses working memory for lexical items by requiring participants to listen to a series of sentences and repeat the last word of each one back to the experimenter. Between each item, they must also decide whether the sentence they heard was true or false. Judging by the results of the study done by Janse & Jesse (2014), we surmise that participants with lower Listening SPAN scores will have poorer working memory, or a lesser ability to keep information online, than those with higher scores, making them less able to use contextual information for the recognition of upcoming words. We also know from a study done by Van Petten, Weckerly, McIsaac, & Kutas (1997) that individuals with low working memory capacity are less sensitive to sentence congruity. Thus, we hypothesize that these participants will have less strong contextually-based predictions, and therefore show an attenuated N400 expectancy effect in response to unexpected items.

Subtract 2 SPAN. The Subtract 2 SPAN task (Salthouse, 1988) assesses working memory for numerical items. We propose that participants with lower Subtract 2 Span scores will show the same attenuation of the N400 expectancy effect predicted in the previous hypothesis. Both measures have been included to explore any potential differences between the two types of working memory (lexical versus numerical).

Methods

ERP STUDY

Participants

Participants included 51 individuals between the ages of 18 and 35 (M = 20.9, SD = 4.45). Six of these participants were excluded from the study following artifact rejection. Every participant had normal or corrected to normal vision, no history of hearing problems or current hearing problems, and no history of head trauma. All participants were right-handed, native American English speakers with no history of psychiatric, learning, or neurological disorders. Each participant provided informed consent per the requirements of the Institutional Review Board of Tufts University and received hourly compensation for their time. For this study, only the 20 participants who completed the second experimental session will be included.

Experimental Stimuli

More extensive details surrounding stimulus creation, characteristics, splicing, and grouping can be found in Storch (2016). No modifications were made for this continuation of the experiment.

Construction of Stimuli. First, 128 highly constraining sentence stems were obtained. Each stem was then paired with both a highly expected single-word ending and an unexpected, but still plausible, single-word ending. Half of these unexpected words were then additionally placed in different low-constraint stems, for a total of 64 low-constraint, unexpected but plausible continuations. This resulted in 64 low-constraining sentences with the same critical word as 64 of the high-constraint unexpected scenarios. In addition, 192 fillers, each with an expected and an unexpected continuation, were then obtained.

This resulted in the stimulus categories described in the Word Expectancy and Disfluency manipulation. For instance, one complete set of stimuli was as follows, with disfluency underlined and unexpected endings in bold typeface (Storch, 2016):

(1a) “He put a clean sheet on the bed.”

(1b) “He put a clean sheet on the **clipboard**.”

(1c) “In his haste, the young man had forgotten to sign the **clipboard**.”

(2a) “The groom took the bride’s hand and placed the ring on her uhh finger.”

(2b) “The groom took the bride’s hand and placed the ring on her uhh **dresser**.”

(2c) “Helen reached up to dust the uhh **dresser**.”

Stimulus Characteristics. As per Storch (2016), the average cloze probability of the expected completions was 85.2%, and the average cloze probability of all of the unexpected completions was 3.28%. Critical words were matched for and selected on a number of phonological properties in order to ensure that any experimental results could be attributed to the intended manipulations. For fillers, the average cloze probability of the expected completions was 63%. Disfluencies immediately preceded the final word of each sentence.

Splicing. Each highly constraining sentence was recorded six times and then spliced together to ensure that each context would be acoustically the same for all items in all conditions, and that the fluent or disfluent part of each item would remain consistent as well. Half of the filler sentences were also spliced to eliminate listener bias from hearing only the critical sentences being spliced.

Lists. Resulting stimuli and fillers were counterbalanced across sets of four lists (expected-disfluent, unexpected-disfluent, expected-fluent, and unexpected-fluent) in a pseudorandomized order. Details of pseudorandomization can be found in Storch (2016). These items were then distributed in such a way that each participant would receive a unique composition and order of stimuli in a manner that still remained counterbalanced.

Reliability. Within the experiment, there were two participant groups. For one group of participants, disfluency preceded unpredictable words 75% of the time. This was considered to be the reliable condition, as the speaker was reliably using disfluencies in the way that listeners might ordinarily expect. For the other group, the disfluency was equally likely to precede both predictable and unpredictable words, creating an unreliable condition as well.

Experimental Procedure

Participants sat in a quiet, dimly-lit room and listened to the stimuli over a set of headphones. Volume was kept at a consistent level for each participant; exceptions were noted when individuals requested that the sentences be played at either a higher or lower volume. The self-paced experiment began with a 500ms fixation cross and the presentation of the first sentence. Participants were asked to focus on this fixation cross while listening to each sentence in order to reduce any head or eye movement, blocking, or blinking. At random points during the experiment, participants were asked to answer a comprehension question about the last sentence played in order to ensure that they were paying full attention. Responses were made using buttons on a video game controller that corresponded to either a “yes” response or a “no” response. Following each sentence, a blink sign appeared, and participants could then decide when to proceed to the next sentence by pressing a button on a video game controller. The results of these comprehension questions will not be discussed in the present study. Each participant heard 192 experimental sentences and 192 filler sentences in total, split up into eight blocks of 48 sentences each.

Following the experiment, participants were given a surprise memory test to assess whether disfluencies affect memory for critical words in each participant group. The results of this memory test will also not be discussed in this paper.

EEG Recording

Before each session began, researchers took 200 10-microvolt calibration pulses to be used in normalization of the amplitude data for each participant. A 29-channel EEG cap was used to collect the data, with additional electrodes placed on both the right and left mastoid area (behind each ear), below the left eye, and beside the right eye. The EEG signal was referenced to the left mastoid. When setting up each participant, researchers ensured that impedance was kept below 5 kΩ for all scalp electrodes, below 2.5 kΩ for each mastoid electrode, and below 10 kΩ for the two eye channels. All exceptions were documented following setup. An Isolated Bioelectric Amplifier System, Model HandW-32/BA (SA Instrumentation San Diego, CA) was used to amplify the EEG signal with a bandpass of 0.01 to 40 Hz; the signal was continuously sampled at 200 Hz by an analog-to-digital converter. The stimuli and behavioral responses were monitored by a researcher to ensure that participants were focused and refraining from blinking during stimulus presentation, and that each sensor remained functional throughout the experiment.

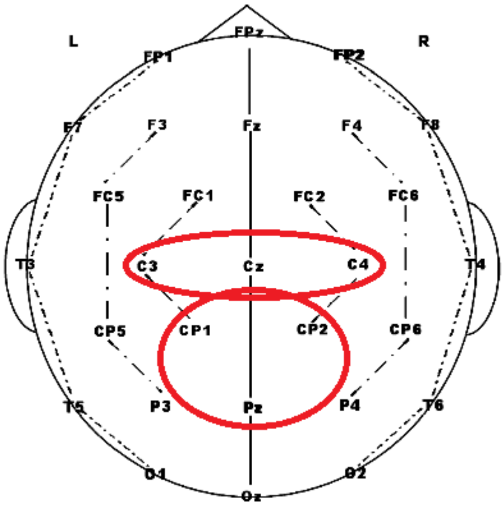
EEG/ERP Processing

All channels were high-pass filtered using a 2nd order Butterworth IIR filter with a half-amplitude high pass cutoff of 0.1Hz, and all epochs were time-locked to stimulus onset and baselined to the average of -100ms to 0ms.[1]

We defined the N400 time window as the average of C3, Cz, C4, CP1, CP2, and Pz electrodes, seen below in red:

---

[1] Thank you to Delaney-Busch & Brown for contributing and phrasing the methodology on this point.



N400 amplitudes for individual trials were trimmed for each subject, such that outliers more than 2.5 standard deviations away from the mean for that subject (about 1.5% of the most extreme trials) were removed. This removes high-leverage trials that can be inferred to carry a low signal-to-noise ratio (for instance, removing this ~1.5% of trials at the Pz electrode reduces the range of N400 amplitudes by 17.5%).[1]

Artifact Rejection

Artifact rejection was performed using the EEGLAB toolbox of MATLAB (Delorme & Makeig, 2004). Researchers first pre-processed the raw data by running it through a script which automatically re-referenced it to the average of the left and right mastoid, filtered it for continuous data (high-pass filter = 0.1), and broke up the data using a window time-locked to the onset of the disfluency (if present) and another window time-locked to the onset of the critical word. Trials were rejected if they contained significant artifact such as blinks, blocking, drift, excessive muscle noise, or horizontal eye movements in a time window from 200ms prior to onset and 900ms after onset for each stimulus. Once the data were pre-processed, researchers used another script responsible for initial detection of blinks, stepwise artifacts, and blocking, as well as detecting peak-to-peak amplitude on all channels. Researchers then went in by hand to change any windows necessary to reject additional artifact and correct any false positives. Mean rejection rate for the 20 subjects included in this analysis was 6.4% (SD = 4.7%).

NEUROPSYCHOLOGICAL BATTERY

Participants

Participants included 20 individuals between the ages of 18 and 35 (M = 21.83, SD = 5.523) who had previously participated in the ERP portion of this study. Each participant provided informed consent per the requirements of the Institutional Review Board of Tufts University and received hourly compensation for their time.

Experimental Procedure

All participants were administered the same battery of neuropsychological tests to assess cognitive function. On average, the testing session lasted two and a quarter hours. A subset of tests relevant to this study are listed below. Of these, the Listening SPAN and Subtract 2 SPAN tasks were administered by the experimenter, while the AQ, AXCPT, SPQ, Stroop Test, and VSL were taken by participants on a desktop computer in the behavioral testing room with the experimenter reading each set of instructions aloud.

Tests administered directly by the experimenter were scored by hand during the session. Tests administered on the computer recorded scores automatically. Each session was also audio recorded up until the start of the computer tasks. Upon completion of the battery, two researchers would score each task based on this recording, after which one additional researcher with no previous exposure to the session would compare the two records for scoring accuracy.

Neuropsychological Tests

AX-Continuous Performance Task. During the AX-Continuous Performance Task (AXCPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956), participants are required to view a continuous string of letters flashing on a screen and respond one way if they see an “A” followed by an “X,” and respond another way if they see any other combination of letters (e.g. “AY,” “BX,” or “CD”). The test was taken on a desktop computer in the behavioral room using two keyboard keys and was automatically scored for accuracy upon completion of the script.

Stroop Test. The Stroop Test (Stroop, 1935) requires participants to respond to stimuli words based on the color in which they are written while ignoring what the words actually say (e.g. pressing the keyboard key corresponding to “blue” when presented with the word “red” in blue typeface). The test was taken on a desktop computer in the behavioral room using four keyboard keys and was automatically scored for speed and accuracy upon completion of the script.

Visual Statistical Learning. The Visual Statistical Learning (VSL; Fiser & Aslin, 2002) consists of an exposure phase and a test phase. During the exposure phase, participants view a series of 288 images that are presented on the computer screen. There are twelve unique images in this series, and they are presented in groups of three. Participants are required to press the spacebar whenever they see an image presented twice in a row. During the test phase, these images are presented in triplets. In each round, participants view a pair of triplets (one they have seen before and a foil triplet that was never presented in the exposure phase) and must decide using two keyboard keys which triplet is more familiar to them. Each triplet is tested eight times, for a total of 32 test trials within this phase.

Autism Spectrum Quotient. The Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) consists of 74 items that participants must respond to with either “definitely agree,” “slightly agree,” “slightly disagree” or “definitely disagree.” The items were designed to assess five different areas commonly considered to be impacted in individuals on the autism spectrum: social skill, attention switching, attention to detail, communication, and imagination. Participants were given this questionnaire on a desktop computer with the knowledge that their responses would remain confidential, ensuring honest answers to each question.

Schizotypal Personality Questionnaire. The Schizotypal Personality Questionnaire (SPQ; Raine, 1991) consists of 74 items that participants must respond to with either “yes” or “no.” The SPQ has been found to have high sampling validity, high internal reliability (0.91), test-retest reliability (0.82), convergent validity (0.59 to 0.81), criterion validity (0.63, 0.68), and discriminant validity (Raine, 1991). Participants were given this questionnaire on a desktop computer with the knowledge that their responses would remain confidential, ensuring honest answers to each question.

Listening SPAN. During the Listening SPAN (Daneman & Carpenter, 1980) task, participants hear a series of sentences played over a set of speakers. They are required to state after each sentence whether it was true or false. After each sequence ends, participants must repeat back the last words of each sentence heard. The sequences start with two sentences and gradually become longer, concluding with sequences of six sentences each. Participants hear fifteen sequences in total.

Subtract 2 SPAN. During the Subtract 2 SPAN (Salthouse, 1988) task, participants hear a sequence of numbers played over a set of speakers, and then repeat that sequence back to the experimenter after first subtracting two from each digit. The sequences start with two numbers each and gradually get longer, concluding with a total of eight numbers. Participants hear 35 sequences in total.

BEHAVIORAL DATA PROCESSING

AXCPT. There were two measures taken from each participant's overall AXCPT performance. First, we calculated the mean RT difference between the AY and the AX trials. We then log-transformed this reaction time difference in order to reduce skew. Second, we calculated a Proactive Behavioral Index to show the degree to which participants were employing proactive strategies (expecting an X when given an A) versus reactive strategies (thinking back to whether the previous trial was an A when given an X). This was calculated as the mean reaction time difference between AY trials (where slowdown indicates proactive control) and BX trials (where slowdown indicates reactive control) divided by their sum (to control for overall speed, where an equivalent difference will yield greater proactive control values in participants with generally fast reaction times).[1] This PBI measure was then log-transformed.

Stroop. The reaction time difference between incongruous and congruous trials for each participant was used as the outcome measure. RT differences were then log-transformed.

VSL. VSL test score was used as the outcome measure.

AQ. Each participant's responses were used to generate a total score by adding up the individual scores for each question that indicated higher AQ.

SPQ. Each participant's responses were used to generate a total score by adding up the number of questions indicating schizotypy with a "yes" response. The results of this summation are not included in the present analysis.

Listening SPAN. Total number of items correct for each participant was used as the outcome measure.

Subtract 2 SPAN. Each participant's SPAN (calculated based on the highest span for which the participant responded to 3/5 items correctly) was used as the outcome measure.

For modeling purposes, all individual differences measures were then z-transformed so that they could be more easily compared to one another.

STATISTICAL ANALYSIS

Hypotheses were tested using multilevel regression modeling, with single-trial N400 amplitudes as the outcome and trial-level (e.g. expected vs. unexpected trial) and subject-level information (e.g. Listening SPAN score) entered as predictors. Each model used the maximal random-effects structure (Barr, 2013) supported by the data (Bates, Kliegl, Vasishth, & Baayen, 2015). Models were fit using restricted maximum likelihood estimation as implemented in the "lme4" package version 1.1-11 (Bates, Mächler, Bolker, & Walker, 2015) in the R statistics software program version 3.2.4 Revised (R Core Team, 2014). Hypothesis tests for individual cross-level interactions were conducted using an F-test of the type-III sums of squares variance explained by the interacting predictors, with denominator degrees of freedom estimated using a Satterthwaite approximation, as implemented by the lmerTest:anova function of the lmerTest package version 2.0-30 (Kuznetsova, Brockhoff, & Christensen, 2015). For tests involving adaptation, "event number" (i.e. ordinal position within the experiment) was entered into the model, accounting for change in the N400 amplitudes and effects over time.[1]

For the majority of tests, we looked at all cross-level interactions. However, for tests focusing on adaptation (such as AQ and VSL), we only generated results relevant to the hypotheses at hand.

Results

ERP STUDY

Within this subset of participants, we found a significant main effect of disfluency (F(1, 29.25) = 10.17, p = 0.0034). We also found a significant main effect of expectancy (F(1, 23.40) = 17.82, p < 0.001). There was no significant interaction between the two (F(1, 260.25) = 0.67, p = 0.413). In addition, there was also no main effect of group, allowing us to remove this manipulation from the following statistical analyses (F(1, 21.60) = 1.03, p = 0.322).

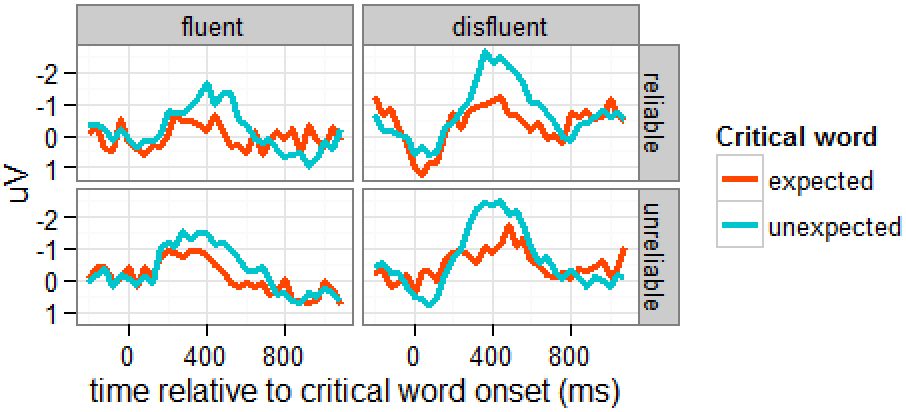


Figure 1. Average N400 amplitude over central and posterior sensors (our ROI) for the reliable and unreliable conditions. There was no significant difference between the N400 effects seen across the two groups.

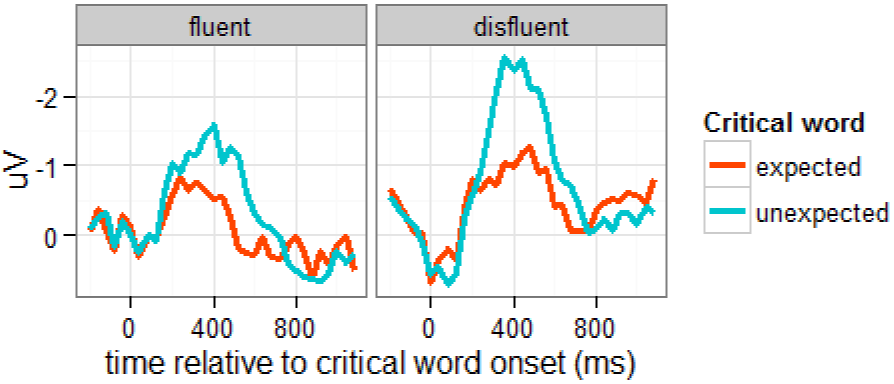


Figure 2. Average N400 amplitude over the same ROI, collapsed across group.

BEHAVIORAL RESULTS

AXCPT. There was no statistically significant effect of AX:AY differentiation on the N400 expectedness effect (F(1, 19.06) = 0.352, p = 0.560). These results do not support our hypothesis that individuals with greater AY-trial reaction times would exhibit an amplification of the N400 expectedness effect.

Additionally, there was no statistically significant effect of the PBI on the N400 expectedness effect (F(1, 19.53) = 2.272, p = 0.148). These results also do not support our hypothesis.

Stroop. There was no statistically significant effect of the Stroop slowdown on the N400 expectedness effect (F(1, 19.28) = 0.178, p = 0.678). These results do not support our hypothesis that individuals with greater Stroop slowdown would show an attenuated N400 expectedness effect regardless of the item’s fluency.

Additionally, there was no statistically significant interaction between Stroop slowdown, disfluency, and the N400 expectedness effect (F(1, 691.51) = 0.265, p = 0.607). These results do not support our hypothesis that when presented with a sentence containing a disfluency followed by an expected continuation, participants with less Stroop slowdown would be better able to suppress the prediction generated by the presence of the disfluency, thus attenuating the N400 expectedness effect.

VSL. Individuals' VSL scores were not found to have any significant effect on adaptation to the speaker's disfluency over the course of the experiment, as indexed by the lack of interaction between disfluency and event number ($F(1, \infty^2) = 0.000$, $p = 1.000$). These results do not support our hypothesis that individuals in the unreliable condition with greater VSL scores would show greater amounts of adaptation, and thus an attenuated N400 expectancy effect, as the experiment went on.

AQ. Individuals' AQ scores were not found to have any significant effect on adaptation to the speaker's disfluency over the course of the experiment, as indexed by the lack of interaction between disfluency, event number, and AQ score ($F(1, \infty) = 0.0053$, $p = 0.942$). These results do not support our hypothesis that individuals with greater AQ scores would show lower amounts of adaptation as the experiment went on.

SPQ. We were unable to analyze SPQ scores. This is currently in progress.

Listening SPAN. Individuals' Listening SPAN scores were not found to have any effect on the N400 expectancy effect ($F(1, 20.548) = 0.322$, $p = 0.577$). These results do not support our hypothesis that individuals with lower Listening SPAN scores would show an attenuated N400 expectancy effect regardless of the item's fluency.

Subtract 2 SPAN. Individuals' Subtract 2 SPAN scores were not found to have any effect on the N400 expectancy effect ($F(1, 20.946) = 0.400$, $p = 0.534$). These results do not

---

[2] This value was used as a result of the model's failure to converge.

support our hypothesis that individuals with lower Subtract 2 SPAN scores would show an attenuated N400 expectancy effect regardless of the item's fluency.

Discussion

GENERAL DISCUSSION

The goal of this study was to use an individual differences approach to explore and possibly explain some of the variation in prediction and adaptation that occurs during language processing. The results of this analysis indicate that any variation in prediction and adaptation found across individuals during Brown et al.'s EEG study of disfluency cannot be explained by the neuropsychological differences tested here. However, these results contradict much of the literature surrounding individual differences in language processing, and thus must be evaluated more carefully before making any generalizations or drawing any final conclusions.

Although we ultimately did not find supporting evidence for any of our eight hypotheses, we do know that past research conducted on individual differences indicates that neuropsychological differences between individuals impact many aspects of cognitive function. For instance, the recent ERP study done by Daltrozzo et al. (2017) determined through a study of the effects of statistical learning on natural language ability in adults that response times to a visual statistical learning task, much like the task we employed here, were related to receptive vocabulary and grammatical ability. Their findings also show through ERPs that the association between statistical learning and receptive vocabulary depended on attention: something directly related to our research question, given our results' support of the attention-orienting hypothesis surrounding disfluency. Notably, the effects of statistical learning on natural language comprehension have been observed in a study conducted by Misyak & Christiansen (2011) as well. Together, the results of these studies lend further credence to the idea that differences in individuals' statistical learning abilities have the potential to impact many aspects of language processing.

Furthering that point, it is possible that the abilities that we assessed might be domain-specific rather than domain-general, as our hypotheses assumed. Although the studies done by Daltrozzo et al. (2017) and Misyak & Christiansen (2011) seem to imply that visual statistical learning abilities impact natural language comprehension, they do not confirm that implicit visual statistical learning translates to linguistic statistical learning. Ultimately, abilities assessed by a task such as our Visual Statistical Learning task, which examines individuals' abilities to implicitly learn image triads, may not transfer to implicit learning of linguistic regularities. It is possible that this may have contributed to our null results.

On the contrary, if it is true within the general public that the neuropsychological measures tested here do *not* have an effect on prediction and adaptation during language comprehension, as our exploratory results suggest, then we might begin to question what *does* cause us to see differences between participants. For instance, perhaps it is not working memory capacity that predicts an individual's comprehension abilities, but rather the quality of their stored phonological representations (MacDonald & Christiansen, 2002). This would attribute individual differences in sentence processing to variation in skill and experience, rather than the limits on storage and computational capacity that might occur as a result of working memory (Just & Carpenter, 1992; Waters & Caplan, 1996).

LIMITATIONS AND OPEN QUESTIONS

In conducting our analyses, there were potentially relevant factors that we did not test. For instance, a number of our participants were bilingual. Bilinguals have been shown in a number of previous studies to have better executive function, performing better on tasks that involve cognitive processes such as problem solving, mental flexibility, attentional control, inhibitory control, and task switching (Marian & Shook, 2012; Bialystok & Craik, 2010; for a comprehensive review, see Adesope, Lavin, Thompson, & Ungerleider, 2010). While we must also consider the recent research indicating that these effects may be due to a publication bias rather than any real significant benefit (Bruin, Treccani, & Sala, 2015), it may be worthwhile to take participants' language-learning history into account when conducting any further analysis. Assuming the bilingual advantage does exist, there might be effects of this higher executive function, and particularly of stronger inhibitory control, on context- or disfluency-based prediction suppression, leading to different neural signatures of language processing within a study such as this one.

Additionally, beyond outside factors, there are unanalyzed aspects of the neuropsychological tasks used in this study that may have had effects on our results as well. For instance, it is notable that many of the items that participants are required to keep online during the Listening SPAN task are from low-constraint scenarios, requiring the listener to generate predictions for sentences with many possible continuations. Often, this will lead to prediction violation. A strong working memory of these unexpected items might mean that the individual can better process these prediction violations and thus better adapt to the nuances of an irregularly disfluent speaker. Many of the tasks that we used have nuances like this one; it is impossible to account for every factor of every one, leaving us with a wealth of questions that we have yet to ask.

One immediate limitation of this study is its sample size. This has been raised as a concern in previous reviews of individual differences analyses of language processing, particularly when using electrophysiological methods, as a small sample size limits the statistical techniques that one can use during data analysis (Boudewyn, 2015). Although we administered the ERP component of this study to 51 people, only 20 of them returned for the neuropsychological testing session. This raises questions of bias within the sample. While many of the original 51 participants were from the greater Boston area, a majority of the participants who returned for the second experimental session were students from Tufts University where the study took place. It can be assumed from the entry requirements of the university itself that all of these participants were of high IQ and high executive function, leading us to question just how representative of our larger sample size, much less the population, our 20 participants were. We may require more variability in order to thoroughly explore the relationships that we sought to study.

Apart from recruiting additional subjects and expanding our population to young adults outside of Tufts, there is also the possibility that diversifying the ages of our participants might impact our results. An EEG study by Federmeier, Kutas, & Schul (2010) lends credence to this idea. Their paradigm involved presenting younger and older adults with phrasal cues for category exemplars (e.g. “An insect”) or antonyms (e.g. “The opposite of closed”), followed by targets that were congruent (“ant” or “open”) with the cues. They also had another condition with congruent but low-typicality category targets (e.g. “hornet” following “An insect”). With the knowledge that there is significant N400 priming for congruous (compared to incongruous) targets in all age groups (Iragui & Kutas, 1993), they sought to examine predictive processing.

Overall, their data showed that young adults predict likely targets when presented with phrasal cues in the context of making acceptability judgments. These predictions are said to preactivate semantic, lexical, and orthographic information associated with likely future words, facilitating processing when those expected targets are encountered. However, when predictions are incorrect, but the input is still plausible and thus must be semantically accepted, they found that additional processing was involved, presumably to suppress or revise the initial prediction and learn from that prediction error (Federmeier et al., 2010). Although these results are more aligned with the “expect the unexpected” hypothesis of disfluency, they do tell us that there is an effect of age on prediction due to differences in the utilization of predictive processing. If older adults make use of language input online differently from younger adults, then we might very well find differences between the ways that they utilize and adapt to disfluency as well.

It is also notable that although the present analysis did not yield any effects of speaker reliability on prediction, analyses of the larger EEG participant base did. This indicates to us that there may be additional differences between the subset of participants used for our purposes and the greater population. This potential effect of reliability also opens the doors for many further research questions, such as whether the reliability of the speaker might impact rates of adaptation, as it indicates that listeners’ implicit knowledge of a speaker plays a part in how they might use the speaker’s disfluency. In the future, we may look to studies done by researchers like Kleinschmidt and Jaeger (2015), whose work examines listeners’ abilities to recognize the familiar, generalize to the similar, and adapt to the novel, to inform our ideas about how listeners respond to speaker variance and delve more deeply into these results.

ERP RESULTS

Within our subset of participants, we did find a statistically significant N400 effect of cloze, as well as a significant N400 effect of fluency. These results show a proof of concept for the classic N400 cloze effect, and also indicate that participants used disfluencies systematically in generating their lexical predictions. In this experiment, N400 amplitude was found to increase when listeners were presented with an unexpected word preceded by a disfluency, which notably contradicts the results found by Storch (2016) in her preliminary dataset. Rather than using disfluency as a cue to expect the unexpected, it seems from our results that the presence of the disfluency might actually be orienting listeners' attention, signaling them to make stronger predictions than they might have when faced with a fluent item. These results support work done in recent years by researchers such as Fraundorf and Watson (2011) as well as Collard, Corley, MacGregor, and Donaldson (2008).

CONCLUSIONS

In sum, while our particular exploratory study did not bring to light any effects of neuropsychological differences on prediction and adaptation, it is clear to us that there still remains work to be done using this approach. There are many directions that we have yet to take our data in, and many participants that we have yet to gather individual differences data from. After all, the implications of a study such as this one are great. We know that people differ, and that these differences affect the way they process and comprehend language. With significant results, we could inform larger fields such as adaptation and pragmatics beyond our present contributions to the disfluency literature.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, 80(2), 207-245. doi:10.3102/0034654310368803

Anand, A., Wales, R. J., Jackson, H. J., & Copolov, D. L. (1994). Linguistic Impairment in Early Psychosis. *The Journal of Nervous and Mental Disease*, 182(9), 488-493. doi:10.1097/00005053-199409000-00002

Anderson, M. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, 49(4), 415-445. doi: 10.1016/j.jml.2003.08.006

Arnold, J. E., Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 914-930. doi:10.1037/0278-7393.33.5.914

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders,* 18(3), 379-402. doi:10.1007/bf02212194

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5-17.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. doi:10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious Mixed Models. *Journal of Memory and Language*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). doi:10.18637/jss.v067.i01

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22(3), 201-211. doi: 10.1177/002383097902200301

Bialystok, E., & Craik, F. I. (2010). Cognitive and Linguistic Processing in the Bilingual Mind. *Current Directions in Psychological Science,* 19(1), 19-23. doi:10.1177/0963721409358571

Boudewyn, M. A. (2015). Individual Differences in Language Processing: Electrophysiological Approaches. *Language and Linguistics Compass*, 9(10), 406-419. doi:10.1111/lnc3.12167

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia*, 50(11), 2659-2668. doi: 10.1016/j.neuropsychologia.2012.07.019

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2013). Effects of working memory span on processing of lexical associations and congruence in spoken discourse. [Original Research]. *Frontiers in Psychology*, 4, 60. doi: 10.3389/fpsyg.2013.00060

Braver, T. S. (2013). Dual Mechanisms of Cognitive Control in AXCPT. Retrieved from http://pages.wustl.edu/dualmechanisms/axcpt-task

Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. *Trends in Cognitive Science*, 16(2), 106-113. doi:10.1016/j.tics.2011.12.010

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, 136, 135-149. doi: 10.1016/j.cognition.2014.10.017

Brown, M., Delaney-Busch, N., Storch, B., Wlotko, E., & Kuperberg, G. (2017). *What's, uhh, coming next? ERP adaptation to distributional characteristics of disfluency*. Abstract from the 30th Annual CUNY Conference on Human Sentence Processing, Massachusetts Institute of Technology, Cambridge, MA

Bruin, A. D., Treccani, B., & Sala, S. D. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science*, 26(1), 99-107. doi:10.1177/0956797614557866

Collard, P., Corley, M., Macgregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(3), 696-702. doi:10.1037/0278-7393.34.3.696

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73-111. doi:10.1016/s0010-0277(02)00017-3

Daltrozzo, J., Emerson, S. N., Deocampo, J., Singh, S., Freggens, M., Branum-Martin, L., & Conway, C. M. (2017). Visual statistical learning is related to natural language ability in adults: An ERP study. *Brain and Language*, 166, 40-51. doi:10.1016/j.bandl.2016.12.005

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450-466.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491-505. doi:10.1111/j.1469-8986.2007. 00531.x

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(3), 149-161. doi: 10.1016/j.bandl.2010.07.006

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458-467. doi:10.1037/0278-7393.28.3.458

Fraundorf, S. H., & Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, 65(2), 161-175. doi:10.1016/j.jml.2011.03.004

Gernsbacher, M.A. (1990). *The structure-building framework: What it is, what it might also be, and why. Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gernsbacher, M.A. (1997). *Two decades of structure building*. Discourse Processes, 23, 265–304.

Gernsbacher, M.A., & Faust, M.E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17(2), 245–262.

Iragui, V. J., Kutas, M., Mitchiner, M. R., & Hillyard, S. A. (1993). Effects of aging on event-related brain potentials and reaction times in an auditory oddball task. *Psychophysiology*, 30(1), 10-22. doi:10.1111/j.1469-8986.1993.tb03200.x

Janse, E., & Jesse, A. (2014). Working memory affects older adults' use of context in spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, 67(9), 1842-1862. doi:10.1080/17470218.2013.879391

Just, M. A. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127(8), 1811-1821. doi:10.1093/brain/awh199

Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99(1),122–149.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148-203. doi:10.1037/a0038695

Kuperberg, G.R., & Caplan, D. Language Dysfunction in Schizophrenia. In R.B. Schiffer, S.M. Rao, and B.S. Fogel (eds): *Neuropsychiatry*, 2nd edition, 2003: 444-66. Lippincott Williams and Wilkins, Philadelphia, PA.

Kuperberg, G. R., & Jaeger, T. F. (2015). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59. doi:10.1080/23273798.2015.1102299

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463-470. doi:10.1016/s1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62(1), 621-647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205. doi:10.1126/science.7350657

Kuznetsova, A., Brockhoff, P.B., & Christensen, R. H. B. (2014). lmerTest: Tests in linear mixed effects models (version 2.0-20). http://CRAN.R-project.org/package=lmerTest

Lesh, T. A., Westphal, A. J., Niendam, T. A., Yoon, J. H., Minzenberg, M. J., Ragland, J. D., . . . Carter, C. S. (2013). Proactive and reactive cognitive control and dorsolateral prefrontal cortex dysfunction in first episode schizophrenia. NeuroImage: Clinical, 2, 590-599. doi:10.1016/j.nicl.2013.04.010

MacDonald, M.C., & Christiansen, M.H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54.

MacGregor, L. J., Corley, M., & Donaldson, D. I. (2010). Listening to the sound of silence: disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, 48(14), 3982-3992. doi: 10.1016/j.neuropsychologia.2010.09.024

Marian, V., & Shook, A. (2012). The Cognitive Benefits of Being Bilingual. *Cerebrum*, 12. Retrieved from http://dana.org/Cerebrum/2012/The_Cognitive_Benefits_of_Being_Bilingual/

Misyak, J. B., Christiansen, M. H. (2012). Statistical Learning and Language: An Individual Differences Study. *Language Learning*, 62(1), 302-331. doi:10.1111/j.1467-9922.2010.00626.x

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. [Research Support, N.I.H., Extramural]. *Journal of Cognitive Neuroscience*, 22(12), 2886-2898. doi: 10.1162/jocn.2009.21400

Perfetti, C. A., E. W. Wlotko, and L. A. Hart. (2005).Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31. (6), 1281.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Raine, A. (1991). The SPQ: A Scale for the Assessment of Schizotypal Personality Based on DSM-III-R Criteria. *Schizophrenia Bulletin*, 17(4), 555-564. doi:10.1093/schbul/17.4.555

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, 20(5), 343-350. doi:10.1037/h0043220

Salthouse, T. A. (1988). The Role of Processing Resources in Cognitive Aging. *Springer Series in Cognitive Development Cognitive Development in Adulthood*, 185-239. doi:10.1007/978-1-4612-3852-2_7

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology*: General, 125(1), 4-27. doi:10.1037//0096-3445.125.1.4

Storch, B. (2016). *Learning to Expect the Unexpected: Speech Disfluencies, Expectancy, and the N400 Effect.* Unpublished senior honors thesis, Tufts University, Medford, MA.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643-662. doi:10.1037/h0054651

Tsai, N., Au, J., & Jaeggi, S. M. (2016). Working memory, language processing, and implications of malleability for second language acquisition. *Cognitive Individual Differences in Second Language Processing and Acquisition Bilingual Processing and Acquisition*, 69-88. doi: 10.1075/bpa.3.04tsa

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working Memory Capacity Dissociates Lexical and Sentential Context Effects. *Psychological Science,* 8(3), 238-242. doi:10.1111/j.1467-9280.1997.tb00418.x

Waters, G.S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, 103(4), 761–772.