November 2021

# An ERP Study of the Effects of Iconic and Nonsense Gestures on Memory Formation

Brianna E. Cairney
*Louisiana State University and Agricultural and Mechanical College*

# AN ERP STUDY OF THE EFFECTS OF ICONIC AND NONSENSE GESTURES ON MEMORY FORMATION

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Brianna Estelle Cairney
B.A. University of Washington, 2016
December 2021

## Table of Contents

## Abstract

Co-speech gesture is an important part of human communication and aids in comprehension, learning, and memory. The addition of iconic gestures to speech has been shown in prior work to enhance memory for the speech. However, it remains unclear as to whether this benefit requires gestures to be meaningful, or, conversely, if any attentionally-engaging gesture will enhance memory. In the current study, we tested two theories to explain the mnemonic benefits of co-speech gesture: Dual Coding Theory, which attributes these benefits to multimodal encoding and enhanced imageability, and Attentional Highlighting Theory, which posits that gestures draw more attention to concurrent speech. We recorded continuous EEG data while participants watched videos of an actor reciting novel word pairs, and tested the effects of adding of an iconic or a nonsense gesture to the first word of each pair on mental imagery, item memory (memory for gestured words), or associative memory (memory for entire pairs). We found that iconic gestures, but not nonsense gestures, enhanced subjective ratings of imageability and improved item memory. However, these benefits did not extend to pair memory, despite an overall association between perceived imageability and pair memory. ERP analyses suggested that the presence of a gesture concurrently with the first word in each pair may have detracted from the processing of the second words, thus limiting the benefits of iconic gestures to item memory. Overall, these results provide tentative support for a dual-coding based explanation of the mnemonic benefits of co-speech gesture.

# Introduction

One important aspect of human communication is gesture. Humans begin gesturing as early as 10 months old (Bates, 1979) and continue to frequently utilize co-speech hand movements throughout life. A substantial amount of research has demonstrated that co-speech gestures enhance communication (Driskell & Radtke, 2003; Goldin-Meadow, 1999; Hostetter, 2011) and facilitate learning and memory (Beattie & Shovelton, 2001; Church, Garber, & Rogalski, 2007; Cook, Yip, & Goldin-Meadow, 2010; Thompson, Driscoll, & Markson, 1998). However, the mechanisms that underlie these benefits remain to be specified. Moreover, it remains unclear as to whether distinct types of gestures affect memory to different degrees and/or via different neurocognitive mechanisms.

Seminal work by McNeill (1992) provides operational definitions of different types of gestures, including iconic and beat gestures. Iconic gestures provide a visual representation of the words with which they are presented, such as raising the hands, palms upward while describing a person *lifting* a box. Beat gestures are small finger or hand movements that do not convey semantic meaning but are thought to emphasize concurrent speech. Each of these types of gestures contribute to communication in meaningful, but often different, ways.

Spontaneous beat gestures can lead to improvements in free and cued recall performance. For example, in Cook, Yip, & Goldin-Meadow (2010), participants who observed and then described short vignettes later recalled more details of those short stories when they produced beat gestures during their descriptions. Observing co-speech beat gesture may similarly enhance memory for verbal information. Participants who watched video clips of a person reciting verbs later recalled significantly more words that were paired with beat or iconic gestures compared to

those spoken without gestures, and there was no difference in recall between the two gesture conditions (So, Sim, & Low, 2012).

The mnemonic benefits of beat gestures may be the result of increased attention to the information being presented. Beat gestures are rapidly integrated with concurrent speech, as evidenced by modulations of very early (before 100 ms) Event Related Potentials (ERPs) for beat-gestured compared to non-gestured words among participants observing an ongoing speech (c; Kelly, Kravitz, & Hopkins, 2004). Beat-gestured words also elicit an enhanced amplitude of the auditory P2 ERP component (Biau & Soto-Faraco, 2013; Dimitrova et al., 2016), which occurs between 200-300 ms post-word onset and is modulated by attention (Luck, Fan, & Hillyard, 1993). These early ERP differences may reflect an "attentional highlighting" function of beat gesture that helps the listener to identify and attend to important information embedded in speech.

The benefit of beat gestures may be explained by what we refer to here as the Attentional Highlighting Theory, which posits that the addition of gesture draws attention to speech (Biau & Soto-Faraco, 2013; McNeill, 1992). Attentional Highlighting Theory suggests that the benefits of gesture are a consequence of the co-occurrence of movement with speech, and not necessarily the content of the gesture. Multimodal cues enhance attention and improve performance during cognitively demanding tasks (Matusz and Eimer, 2011; Santangelo and Spence, 2007). Gestures and co-speech are a form of multimodal communication, which may increase the saliency of emphasized parts of discourse and capture attention more effectively compared to unimodal communication (i.e., non-gestured parts of speech). The addition of gesture may make verbal discourse more dynamic, resulting in increased attention and subsequent improvements in memory performance.

Like beat gestures, details of speech presented with iconic gestures are recalled more accurately than when speech is presented without any gestures (Beattie & Shovelton, 2001; Thompson, Driscoll, & Markson, 1998). However, the mechanisms by which iconic gestures enhance memory may partially or completely differ from those pertaining to beat gestures. One key difference between beat and iconic gestures is the latter provides semantic support, as evidenced by numerous ERP studies measuring the N400 ERP, a robust indicator of the relative ease or difficulty of semantic processing (Kelly, Kravitz, & Hopkins, 2004; Kutas & Hillyard, 1984; Ozyurek et al., 2007; Wu & Coulson, 2007). Although it is intuitive to suppose that semantic reinforcement contributes to the mnemonic benefits associated with iconic gestures, the underlying mechanisms are not obvious. For example, iconic gestures may help to disambiguate speech (Holle & Gunter, 2007) but it does not necessarily follow that making critical words easier to process should lead to enhanced memory for non-ambiguous speech. There thus exists a lack of specificity in theories of how semantically congruent gestures enhance memory.

A different process that may contribute to iconic gesture-driven memory improvements but has received little attention in gesture research is mental imagery. According to Paivio's Dual Coding theory, memory for verbal information is strongest when it is presented with synchronous visual information (Paivio, 1969; Paivio & Csapo, 1973; Yates, 1966), including mental imagery. Dual coding can be thought of as an integrated process in which verbal information is encoded as a verbal representation; visual information is encoded as a visual representation; and—critically—referential connections are encoded linking the two together (Mayer & Anderson, 1991). Dual-coding theory posits that the well-established memory advantages for concrete (e.g., *apple, truck*) relative to abstract words (*courage, idea*), can be explained by the greater ability of concrete words to evoke mental images, which allows them to

support dual coding. Viewed through the lens of dual coding theory, the addition of meaningful iconic gestures to speech may similarly enhance imageability by providing the listener with a visual representation of the spoken content, effectively "lending" it additional concreteness.

Another key tenet of Dual Coding Theory is that the benefits of imageability on memory can extend beyond the imageable words themselves by facilitating the formation of associations in memory with other stimuli. For example, in a study conducted by Paivio (1965), participants listened to a list of unrelated noun-noun word pairs presented as either concrete-concrete (e.g., *apple-truck*), concrete-abstract (e.g., *apple-courage*), abstract-concrete (e.g., *courage-apple*), or abstract-abstract (e.g., *courage-idea*). Cued recall was higher for pairs with concrete first nouns, regardless of whether the second noun was concrete or abstract. Said differently, words that permit dual coding have the ability to serve as "conceptual pegs" that other words can "hang" onto. Thus, to the extent that iconic gestures serve as "concreteness enhancers", we should similarly predict that they would facilitate both item and associative memory.

The results of a study by Cairney et al. (in preparation) provide preliminary support each of these predictions. In this study, participants watched videos of an actor reciting sentences that ended with unrelated word pairs (e.g., *She thought about the narrowing turtle*). The first word (W1) in each sentence was presented with either an iconic gesture, a beat gesture, or no gesture, and was followed by a semantically unrelated noun (W2) that did not have any accompanying gesture. Using a 6-point Likert-type scale, participants rated how easy or difficult it was to create a mental image of the concept denoted by each word pair. After each block, participants completed a free recall task in which they were asked to write down as many of the word pairs as they could remember. Word pairs with iconic-gestured W1s were recalled at higher rates than pairs for which the W1s were presented with beat or no gestures, whereas no difference in recall

performance was found between pairs with beat-gestured and non-gestured W1s. Iconic-gestured word pairs were also rated as easier to imagine than non-gestured pairs, and marginally easier to imagine than those in the beat gesture condition.

Taken together, the results of Cairney et al. (in preparation) are consistent with a dual-coding view of the benefits of co-speech iconic gesture on memory. However, a potential limitation of this study was that the same beat gesture was used for all beat-gestured word pairs. As a result, the beat gestures may not have captured the participants' attention as much as the unique iconic gestures did, raising the possibility that enhanced mental imagery did not exclusively account for the memory benefits found for iconic over beat gestures. If the iconic gestures were more engaging, then Attentional Highlighting Theory may also explain these results, particularly insofar as the attention engaged by the gestured W1 also extended to the paired W2.

To distinguish among these possibilities, the current study replicated Cairney et al.'s (in preparation) experiment but used word pairs in which the W1s were presented along with iconic gestures, nonsense gestures, or no gestures. In the current study, we aimed to address which theoretical explanation—Dual Coding or Attentional Highlighting—might explain the mnemonic benefits of co-speech gesture when ambiguous hand movements were introduced. Unlike the beat gesture used in our previous work, nonsense gestures were developed to have similar form, range of motion, and attentional engagement as iconic gestures, while conveying minimal semantic meaning. If gains in memory performance rely on dual coding provided by iconic gestures, then both item and associative recall for iconic-gestured word pairs should be higher than words presented with nonsense gestures. However, if complex gestures improve memory by

capturing the observer's attention, recall for word pairs presented with nonsense gesture should be the same as those presented with iconic gesture.

It is important to consider that, even if recall performance improves in both iconic and nonsense conditions, different types of gesture might support associative memory through different neural mechanisms (e.g., Straube et al, 2014). For this reason, we recorded ERPs concurrently with encoding to gain insight into the neurocognitive processes engaged in response to word pairs with iconic, nonsense, and no gestures. Although recordings lasted for the entirety of each trial, all electrophysiological analyses were limited to W2s to control for potential confounds of sensorimotor processing elicited by the gestured W1s.

We were specifically interested in two ERP components—the P3, a robust component elicited during a variety of attentional tasks, and the N700, which is thought to reflect mental imagery. The P3 typically occurs around 250-500 ms (Donchin, Ritter, & McCallum, 1978; Fabiani, Gratton, Karis, & Donchin, 1987; Polich, 2007) and has been linked to attentional orienting to unexpected or surprising stimuli (Donchin, 1981, Mars et al., 2008). P3 effects are more likely to occur when a surprising stimulus is improbable (e.g., Hirshman, Whelley, & Palij, 1989). For example, Courchesne, Hillyard, and Galambos (1975) had participants visually attend to a fixation dot while numbers flashed on the screen. The number '2' appeared on the screen 90% of the time (i.e., the 'standard' stimulus) while the number '4' only appeared 10% of the time (i.e., the 'deviant' stimulus). The deviant stimulus, '4', elicited increased P3 amplitudes compared to the standard '2'. However, others contest the notion that a stimulus must be infrequent to elicit a P3 response. For example, images subjectively perceived as surprising have been shown to elicit an increased P3 amplitude compared to images rated as unsurprising, despite a similar frequency of presentation (Neville et al., 1982).

The N700 is a frontally distributed late slow wave occurring between 500 and 1000 ms and is thought to reflect mental imagery (Barber et al., 2013; West & Holcomb, 2000). Concrete words elicit bigger amplitudes than do abstract words, and this effect is more pronounced during mental imagery tasks (Gullick et al., 2013). N700 concreteness effects have also been found at the compositional level, though with onsets closer to 700 ms (Lucas et al., 2017). If iconic gestures provide additional concreteness to co-speech and thus help recruit mental imagery processes, they may elicit increased N700 amplitudes compared to non-gestured speech.

In summary, if the memory benefits of co-speech gesture depend on enhanced mental imagery, then the Dual Coding Theory predicts that: 1) more words and word pairs in the iconic gesture condition will be recalled than those in nonsense or no gesture conditions, 2) trial-by-trial imageability ratings will be higher for word pairs containing iconic gestures versus nonsense or no gestures, 3) the P3, an ERP component related to attention, will have a similar amplitudes for the W2s across all three gesture conditions and 4) N700 ERPs to the W2s following iconic-gestured W1s will be greater (more negative) than those that follow nonsense or no gestured W1s. By contrast, if iconic and nonsense gestures improve memory by capturing attention, the Attentional Highlighting Theory predicts that: 1) recall in the iconic and nonsense conditions will be similar to one another and higher than recall in the no gesture condition, 2) imageability ratings will not differ between gesture conditions, and 3) gestures will modulate the P3 but not the N700. Specifically, Attentional Highlighting Theory predicts that P3 amplitudes to W2s following iconic- and nonsense-gestured W1s will be greater than those following non-gestured W1s, indicating increased visual attention to gesture. Alternatively, the W2s of nonsense-gestured pairs may elicit a greater (more positive) P3 amplitude compared to iconic gestures because the gesture content is surprising.

Finally, it bears mention that if W2 P3 amplitudes in the iconic and nonsense gesture conditions are similar, it might be because the probability of seeing nonsense, iconic, or no gestures is equal, in which case, P3 amplitude may not adequately explain attentional orienting effects of nonsense gestures. Consideration should therefore also be given to P3 latency, which reflects stimulus processing and categorization (Folstein & Van Petten, 2011; Kutas, McCarthy, & Donchin, 1977). Surprising stimuli may take longer to identify, as evidenced by previous findings of both slower reaction times and increased P3 latency for surprising compared to expected or normal stimuli (Duncan-Johnson, 1981; Duncan-Johnson & Donchin, 1982; Polich & Donchin, 1988). We reasoned that, because nonsense gestures might be more surprising and take longer to process than iconic gestures, word pairs that begin with nonsense gestures may also evoke longer P3 latencies relative to word pairs that begin with iconic gestures.

## Materials and Methods

### Participants

Thirty undergraduate students (25 female, mean age = 21.3 years, range = 18–28 years) at Louisiana State University participated in the study and were compensated for their time with course credit or payment ($12/hour). A power analysis using G*Power 3.1 software (Faul, Erdfelder, Buchner & Lang, 2009) estimating a power of 0.9 and a small effect size F = 0.27 indicated a minimum sample size of 31[1]. Four additional individuals completed the experiment but were excluded from analyses due to technical problems (*n* = 1), poor EEG data quality (*n* = 2), or learning a first language other than English (*n* = 1). All participants included in the analyses were right-handed, and English was their first and primary language.

### Stimuli

*Iconic and nonsense gesture norming*

One hundred twenty-seven iconic gestures and 57 nonsense gestures were normed on Amazon Mechanical Turk (MTurk) to ensure that iconic gestures resembled the words with which they were paired while nonsense gestures were perceived as nonsensical. Participants (*n* = 60; mean age = 38; age range = 21-79) watched successive video clips of randomly presented iconic and nonsense gestures and rated how meaningful they found the gestures to be on a 5-item Likert-type scale (1 = extremely meaningful, 2 = very meaningful, 3 = moderately meaningful, 4 = slightly meaningful, 5 = not meaningful at all). Then participants responded to the question: "If you HAD to choose one word to describe this gesture, what would it be? Please limit your response to one word." Fifteen iconic gestures previously normed and highly rated as resembling

---

[1] The effect size was derived from Cairney et al. (in preparation), which found a $\eta^2_{partial}$ effect of 0.27. When appropriately converted in G*Power 3.1, the f effect size is 0.608 and the minimum sample size recommended is 20. Due to this computational error, we recruited more participants than necessary.

a specific word (e.g., clapping, cutting, driving) were used as attention checks. Mean meaningfulness scores were calculated for every gesture. Gestures that received a score of 1-2.5 were categorized as iconic, whereas gestures that scored 3.5-5 were categorized as nonsense. Gestures that received a score of 2.51-3.49 were separately reviewed by two experimenters. Of those, consistent responses (i.e., more than 50% of participants provided the same word) were categorized as iconic and inconsistent responses (i.e., more than 50% of open-response words were different) were categorized as nonsense. Two separate experimenters further reviewed the open responses, removed any redundant gestures that looked alike (e.g., *push*, *mow*, *pram*), and identified words that were synonymous but had different gestures (e.g., *fighting* and *hitting*; *wringing*, *clenching*, and *crushing*). Gestures were then divided into three counterbalance groups that did not differ significantly on meaningfulness rating scores, contain synonymous words, or contain similar gestures. Thus, gesture norming resulted in 108 iconic gestures and 36 nonsense gestures. Overall, the average meaningfulness score for the iconic gestures was 2.3 (range 1.2 to 3.4), and the average score for the nonsense gestures was 3.6 (range 3.1 to 4.6).

*Sentence stimuli*

Trials consisted of neutral sentences ending in semantically unrelated word pairs (e.g., *They noticed the locking insect*). We refer to the first word in the pair (e.g., *locking*) as W1 and the second word (e.g., *insect*) as W2. Versions of each sentence were created for all three gesture conditions (iconic, nonsense, and no gesture), which were identical except for the type of gesture that co-occurred with the W1. Video was filmed on a 13-inch MacBook Pro using Photo Booth and edited using Adobe Audition Premier. Audio was recorded in a sound-proof booth using a Marantz Professional MPM-2000 large diaphragm condenser microphone and Marantz Professional solid state recorder (PMD661 MK II). The audio recordings were cropped and

edited using Praat. Timing was standardized such that each trial was 8500 ms long: 2500 ms containing the beginning of the sentence followed by a pause (e.g., "*They noticed…*"); 2000 ms containing the W1 (e.g., "*the locking…"*) and corresponding iconic, nonsense, or no gesture; a 1500 ms pause with the actor's hands still in her lap; and 2500 ms for the W2 (the second word with no gesture; e.g., "*insect…*"; Figure 1). Each sentence had iconic, nonsense, and no gesture versions recorded, and the assignment of gesture type to sentence were counterbalanced across participants (Figure 2).

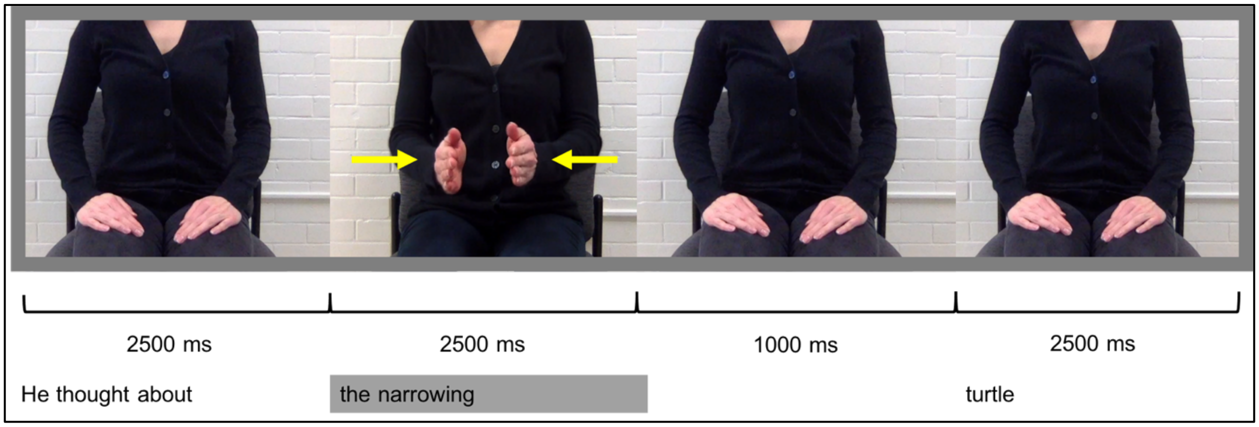


Figure 1. Sample study trial with timing information. Each trial consisted of a video that was edited to be temporally precise and last 8500 ms. During the first 2500 ms, the actor's hands are still while the first part of the sentence is spoken. The hand gesture is edited to occur during the next 2500 ms, beginning in synchrony with the W1. A 1000 ms pause follows, with the actor's hands returned to her lap. The final 2500 ms begin with the W2 followed by stillness and silence.

| Counterbalance Group 1 | Counterbalance Group 2 | Counterbalance Group 3 |
|---|---|---|
| 1 – 36: Iconic | 1 – 36: Nonsense | 1 – 36: None |
| 37 – 72: Nonsense | 37 – 72: None | 37 – 72: Iconic |
| 73 – 108: None | 73 – 108: Iconic | 73 – 108: Nonsense |

Figure 2. Counterbalancing structure of the experiment. Each of the 108 sentences had Iconic, Nonsense, and No Gesture ('None') versions recorded and counterbalanced across three groups. Trials were randomly presented in six blocks, and each block contained 6 Iconic, 6 Nonsense, and 6 None gesture trials.

**Procedure**

The experiment consisted of six 18-trial study blocks, each followed by a memory test. Each block contained 6 iconic, 6 nonsense, and 6 no gesture videos, randomly presented and bounded by one primacy buffer and one recency buffer. Buffer trials were not included in analyses. Data were collected in an acoustically and electromagnetically shielded recording booth. Videos were presented on an ASUS VG248 HD 1920x1080 144 Hz monitor, and audio was delivered through Etymotics E-1 insert earphones.

Participants were informed that their task was to watch the videos and attempt to remember as many word pairs as possible. Participants were further instructed to generate mental images of the concept formed by the word pair. For example, for the pair *locking insect*, the participant might picture an insect locking a padlock with a key. After every trial, participants used a Cedrus button response pad (RB 840, Cedrus, San Pedro, CA) to report, on a 1-6 scale, how easy or difficult it was to generate the image. Half of the participants were told to assign a rating of 1 to word pairs that were the most difficult to image, and to assign a rating of 6 to pairs that were the easiest to image. The mapping of buttons to levels of difficulty was reversed for the remaining participants and re-coded prior to analysis.

Immediately following each block, participants completed a 60 second distractor task in which they counted backwards by twos. After the distractor task, participants were given a sheet of paper and a pen and were instructed to write down as many of the word pairs as they could remember. Participants were told to try their best to recall word pairs, but if they could only remember one word, to still write it down. There was no time limit on word recall.

Participants completed a short practice study-test block before beginning the experiment. The practice block consisted of 4 iconic gesture videos, 4 nonsense gesture videos, and 4 no gesture videos (randomly presented) followed by a practice free recall.

**Electrophysiology**

Continuous EEG data were collected from 32 scalp electrodes throughout the study blocks. An electrode placed on the left mastoid served as an online reference, and data were re-referenced offline to the average of the left and right mastoids. Additional electrodes were placed next to the outer canthus of each eye and below the center of each eye to monitor eye movement (e.g., blinks, saccades). Electrode impedances were monitored throughout the session and kept below 5 kΩ. Electrical signal was recorded using an online bandpass filter of .01-100 Hz and sampled at a rate of 1000 Hz. A bandpass filter of 0.1-30 Hz was applied offline prior to statistical analyses. Epochs were time-locked to the onset of the W2s of each sentence with time windows of -200 to 1000 ms. The 200 ms prior to stimulus onset was used for baseline correction.

Artifact detection was conducted to identify artifacts such as eye movements, blinks, and muscle activity. Simple threshold artifact detection was conducted on 32 scalp electrodes, rejecting trials that contained voltages exceeding $\pm$80 µV. Step-like artifact detection was performed on bipolar vertical and horizontal eye channels with a rejection threshold of $\pm$60 µV, 400 ms windows, and 10 ms window steps. Independent component analysis (ICA) was conducted on datasets that had more than 25% of epochs rejected due to artifacts ($n$ = 9). ICA decomposition was performed on epoched data (excluding ocular channels) using the runica algorithm as implemented in EEGLAB (Delorme and Makeig, 2004). Eyeblink components were manually identified and removed. For datasets with less than 25% of trials rejected for artifacts,

blink trials were excluded from analyses and no ICA was performed ($n$ = 21). Trials containing artifacts due to saccades or muscle activity were eliminated from all data. In total, an average of 8.3% of trials (range = 0–24%) were excluded from each dataset.

Our analysis strategy for the ERPs was three-fold. First, grand-averaged waveforms of P3 and N700 components were generated for each condition (iconic, nonsense, or no gesture). We used a "collapsed localizer" procedure (Luck & Gaspelin, 2017) to select an electrode and time window for the P3. This is an unbiased procedure in which waveforms are averaged across all conditions and groups and the electrodes in which the components of interest are most pronounced are identified from the collapsed average. This resulted in the selection of electrode Fz and a time window of 280-380 ms, (centered around a peak of 330 ms). For the N700, we used an *a priori* 700-1000 ms time window over a frontocentral midline cluster containing electrodes AFz, Fp1, Fp2, Fz, F3, F4, Fc1, Fc2, and Cz (Lucas et al, 2017). Within-subject repeated measures ANOVAs (and follow-up *t*-tests when appropriate) were then conducted on the mean amplitude values between gesture conditions for these components. Our second analysis approach involved the use of Mass Univariate Analysis (MUA) with cluster-based permutation tests, a data-driven approach (described below) which identifies clusters of significant time points and electrode sites while correcting for multiple comparisons. Finally, we conducted exploratory analyses based on visual inspection of the waveforms, which revealed a frontal effect from 600-800 ms that resembled an abbreviated version of hypothesized N700 effect. An additional 10 Hz low-pass filter was applied to grand average waveforms for display purposes only. Data preprocessing was conducted using EEGLAB and ERPLAB (Lopez-Calderon & Luck, 2014), and statistical analyses were conducted using RStudio (RStudio Team, 2021).

When conducting MUA with cluster-based permutation tests, statistical tests are performed on a large number of time points and electrode sites, and a permutation-based cluster mass test serves to control the family wise error rate (Groppe, Urbach, & Kutas, 2011; Fields & Kuperberg, 2020). These analyses were implemented using the Mass Univariate ERP Toolbox (Groppe et al., 2011) and Factorial Mass Univariate ERP Toolbox (Fields, 2017). Data were down-sampled from 1000 Hz to 10 Hz prior to MUA, creating 100 ms time windows. We first conducted a repeated measures ANOVA on ERPs elicited by the W2s, in which W1 gesture (Iconic/Nonsense/None) was the sole factor. This test included all 32 scalp channels and all 100 ms time windows from 0-1000 ms. We used the FclustGND function of the Factorial Mass Univariate ERP Toolbox to identify spatiotemporal clusters by calculating F-statistics for each electrode and time bin using the original data and 10,000 within-subject permutations. Electrodes within 5.44 cm of each other were considered spatial neighbors, and adjacent time points were considered temporal neighbors. Neighboring F-statistics with uncorrected *p*-values of $\leq 0.05$ were grouped into clusters, and the F-statistics within each cluster were summed together to calculate the cluster mass. To assign a *p*-value to each cluster, the cluster masses of the observed data are compared to an estimate of the null distribution based on the largest cluster mass of each within-subject permutation. Clusters showing a significant main effect in this omnibus ANOVA were followed up with pairwise comparisons between conditions. These follow-up tests used similar parameters but were restricted to the time windows and electrodes that comprised the significant clusters.

## Results: Behavior

### Imagery ratings

Figure 3 shows mean imagery ratings for word pairs for each gesture condition. A repeated measures within-subjects ANOVA revealed a significant main effect of W1 gesture on ease-of-imageability ratings, [$F(2, 58) = 7.03$, $p = 0.001$, $\eta_p^2 = 0.2$]. Follow-up paired *t*-tests revealed that participants rated word pairs presented with iconic gestures as easier to imagine than word pairs presented with nonsense gestures, [$t(29) = 2.87$, $p = 0.008$, *Cohen's d* = 0.45] or no gestures, [$t(29) = 2.88$, $p = 0.007$, *Cohen's d* = 0.51]. There was no difference for imagery ratings between the nonsense and no gesture conditions, [$t(29) = 0.84$, $p = 0.41$, *Cohen's d* = 0.08].

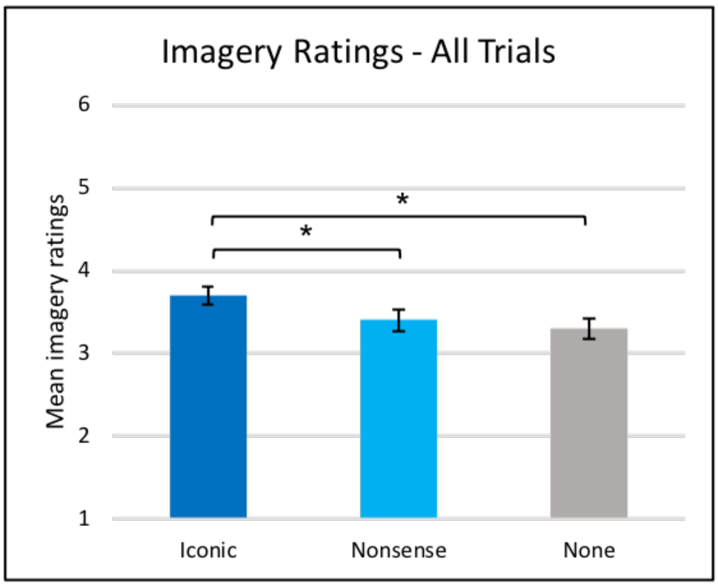


Figure 3. Mean imagery ratings for all word pairs. Participants rated word pairs on a scale of 1 (difficult to imagine) to 6 (easy to imagine). This graph displays the mean imagery ratings for all word pairs, regardless of whether the W1, W2, or entire pair was recalled or forgotten. Word pairs presented in the iconic condition were rated as easier to imagine than word pairs presented in the nonsense and no gesture conditions. There was no difference in reported ease-of-imagery between nonsense and no gesture conditions.

**Free recall**

Mean number of recalled W1s, W2s, and completed pairs are illustrated in Figure 4. To be counted as a completed pair, the W1 must have immediately proceeded the associated W2. If a W1 was remembered but its associated W2 was recalled in a later block as a single response, they were not counted as a completed pair. A 3 (W1 Gesture: Iconic/Beat/None) × 2 (Word: W1/W2) repeated measures ANOVA for word recall was conducted. A significant W1 gesture × word interaction emerged [$F(2, 58) = 3.6$, $p = 0.03$] as well as a significant main effect of W1 gesture [$F(2, 58) = 3.8$, $p = 0.03$], indicating that W1 gesture influenced recall for either W1s or W2s. The main effect of word was not significant [$F(1, 29) = 0.05$, $p = 0.83$].

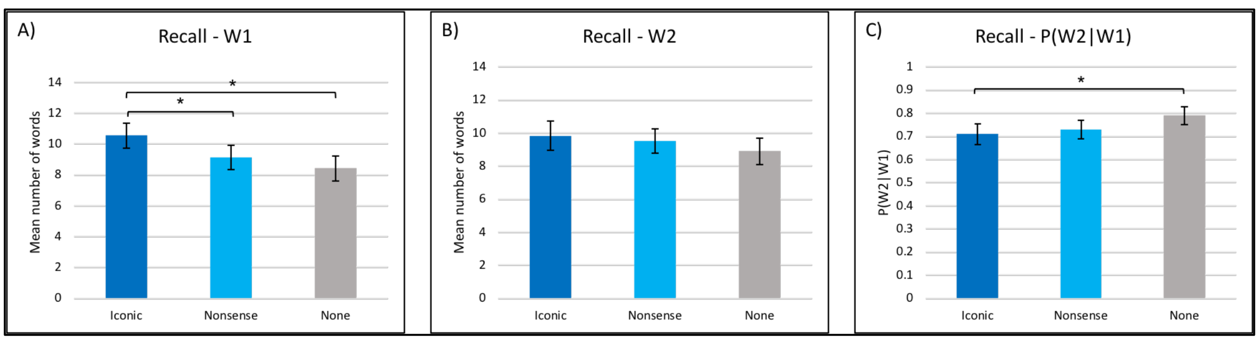


Figure 4. Word recall by condition and recall type. A) Iconic-gestured singleton W1s were recalled at significantly higher rates than nonsense- or non-gestured singleton W1s. There was no difference in recall for singleton W1s between nonsense and no gesture conditions. B) There was no difference in recall for singleton W2s across all three conditions. C) W2s were more likely to be recalled if the corresponding W1 was remembered when the word pair was presented without any gesture compared to word pairs presented with iconic gestures. There was a non-significant trend in conditional probability recall between iconic and nonsense gesture conditions.

To follow up on the significant interaction, two one-way within-subjects repeated measures ANOVAs were conducted to examine the effect of gesture on W1s and W2s separately. The main effect of gesture on recall for W1s was significant [$F(2, 58) = 5.99$, $p = 0.004$, $\eta_p^2 = 0.17$]. As illustrated in Figure 4a, follow-up paired *t*-tests revealed that W1s presented with iconic gestures were recalled more often than W1s presented with nonsense

gestures [$t(29) = 2.22$, $p = 0.03$, *Cohen's d* = 0.33], or no gestures, [$t(29) = 3.3$, $p = 0.003$, *Cohen's d* = 0.48]. There was no significant difference in W1 recall between nonsense and no gesture conditions [$t(29) = 1.18$, $p = 0.25$, *Cohen's d* = 0.16]. No significant effects of gesture emerged on W2 recall [$F(2, 58) = 1.28$, $p = 0.28$, $\eta_p^2 = 0.04$] (Figure 4b).

In summary, these analyses indicate that concurrent iconic gestures, but not nonsense gestures, increased the likelihood that words would be recalled (e.g., item memory). However, unlike in Cairney et al. (in preparation), this benefit did not extend to the paired W2s, suggesting that associative memory was not improved. To more directly test for gesture effects on associative memory, we conducted a within subjects repeated measures ANOVA that examined specifically whether gesture influenced the probability of W2 recall when its associated W1 was correctly remembered (e.g., the conditional probability of W2|W1). The effect of gesture was significant [$F(2, 58) = 3.25$, $p = 0.04$, $\eta_p^2 = 0.1$]. However, follow-up paired *t*-tests revealed that this effect was driven by a significantly *lower* conditional probability in the iconic gesture condition relative to the no gesture condition [$t(29) = 2.31$, $p = 0.03$, *Cohen's d* = 0.38] (Figure 4c). No significant difference was found between the iconic and nonsense gesture conditions [$t(29) = 0.74$, $p = 0.46$, *Cohen's d* = 0.1] nor between the nonsense and no gesture conditions [$t(29) = 1.71$, $p = 0.1$, *Cohen's d* = 0.3]. These analyses provide additional confirmation that W1 iconic gestures had the effect of increasing recall of the W1s themselves but did not aid in the formation of associations of the W1s to their paired W2s. In summary, iconic gestures selectively increased the likelihood of W1s being recalled as single words.

**Imagery ratings and free recall**

Thus far, our results partially support a Dual Coding account of the effects of gesture on memory. Specifically, both W1 recall and subjective imageability ratings were impacted by W1

iconic gestures, but not W1 nonsense gestures. However, unlike in our prior experiment (Cairney et al., in preparation), we did not find evidence that W1 iconic gestures enhanced associative memory. Given the proposed role of enhanced imageability in mediating associative memory improvements, we conducted a final analysis of whether word pairs that were rated as more easily imagined tended to be better remembered. This analysis took the form of a repeated-measures ANOVA with recall type (W1 Only/W2 Only/Pair/None) as the single factor and the mean imageability rating as the dependent variable. The effect of ease-of-imagery on memory was significant [$F(2, 84) = 3.92$, $p = 0.01$, $\eta_p^2 = 0.12$].[2]

Follow-up paired *t*-tests revealed that enhanced mental imagery was greater for correctly recalled word pairs compared to singleton W1s [$t(28) = 2.9$, $p = 0.007$, *Cohen's d* = 0.44], singleton W2s [$t(28) = 2.31$, $p = 0.03$, *Cohen's d* = 0.51], and forgotten trials [$t(28) = 6.1$, $p < 0.001$, *Cohen's d* = 0.76]. There were no significant differences in imagery ratings between singleton W1s and W2s [$t(28) = -0.02$, $p = 0.98$, *Cohen's d* = -0.006], singleton W1s and forgotten trials [$t(28) = 0.56$, $p = 0.58$, *Cohen's d* = 0.09], or singleton W2s and forgotten trials [$t(28) = 0.56$, $p = 0.58$, *Cohen's d* = 0.11] (Figure 5). In sum, these results raise the possibility that iconic gestures may have had an indirect facilitative effect on associative memory by way of enhancing pair imageability that was not strong enough to manifest as a direct effect.

[2] One subject was excluded from the imagery rating × free recall analysis due to insufficient data for one of the conditions.

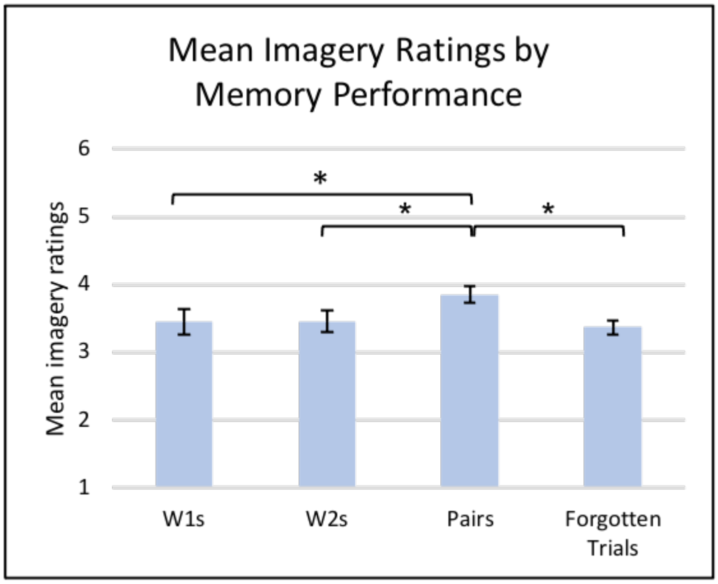

Figure 5. Mean imagery ratings by memory performance. Correctly recalled word pairs had higher reported imagery ratings compared to recalled singleton W1s, recalled singleton W2s, and forgotten word pairs. There was no difference in imagery ratings for recalled singleton W1s, recalled singleton W2s, or forgotten word pairs.

## Results: Electrophysiology

### Analyses based on spatiotemporal averaging

*P3 amplitude*

Using the collapsed localizer approach, we selected the mean amplitude from 280-380 ms at electrode Fz to quantify P3 amplitude (Figures 6a and 6b). A within-subjects repeated measures ANOVA revealed that there was no main effect of gesture on ERPs [$F(2, 58) = 0.63$, $p = 0.53$, $\eta_p^2 = 0.02$].

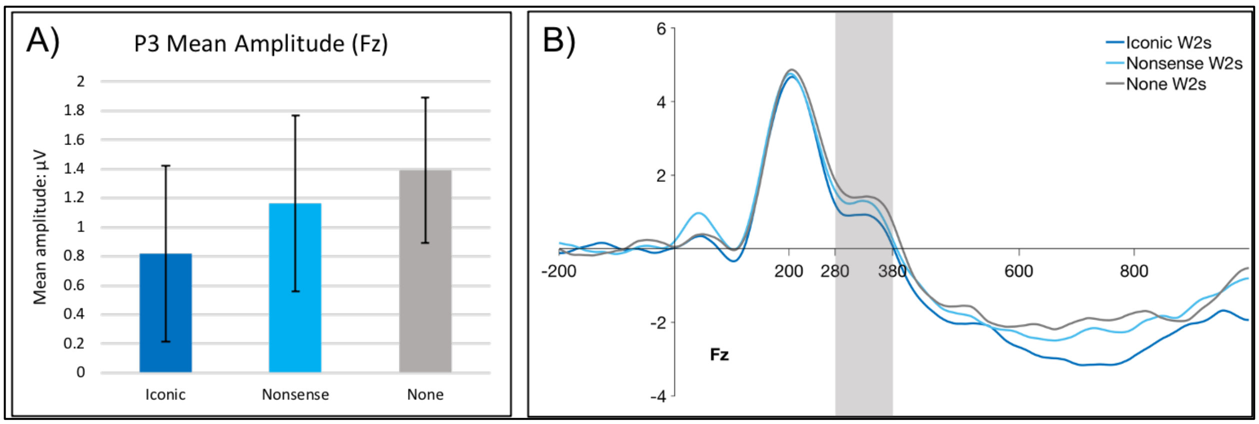


Figure 6. Mean P3 amplitude by gesture condition. The effect of gesture on P3 mean amplitude at electrode Fz was not significant. A) Plotted means of P3 amplitude for all three conditions, which did not significantly differ from one another. B) An ERP waveform of the mean amplitude for all three conditions, with time plotted on the x-axis, voltage plotted on the y-axis, and the measured time window (280-380 ms) shaded.

*P3 latency*

Peak latency at electrode Cz was measured from 200-500 ms (Figures 7a and 7b). A within-subjects repeated measures ANOVA revealed that there was no main effect of gesture on ERP latency [$F(2, 58) = 0.78$, $p = 0.46$, $\eta_p^2 = 0.03$]. This suggests that word pairs were attended to and categorized similarly, regardless of the accompanying gesture (or lack thereof).

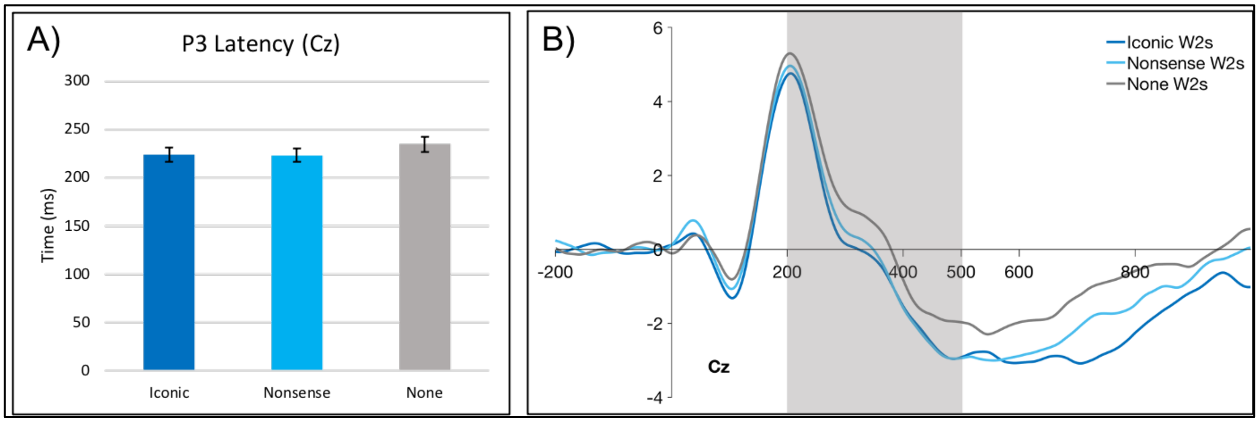


Figure 7. P3 peak latency by gesture condition. The effect of gesture on P3 peak latency at electrode Cz was not significant. A) Mean P3 peak latency values are plotted for each condition, which were not significantly different from one another. B) An ERP waveform of the peak latency for all three conditions, with time plotted on the x-axis, voltage plotted on the y-axis, and the measured time window (200-500 ms) shaded.

*N700 amplitude*

Mean amplitude from 700 to 1000 ms was measured at a frontocentral midline cluster of electrodes (AFz, FP1, FP2, Fz, F3, F4, FC1, FC2, and Cz; Figures 8a and 8b). A within-subjects repeated measures ANOVA was conducted revealed that there was no significant effect of gesture [$F(2, 58) = 2.17$, $p = 0.12$, $\eta_p^2 = 0.07$].

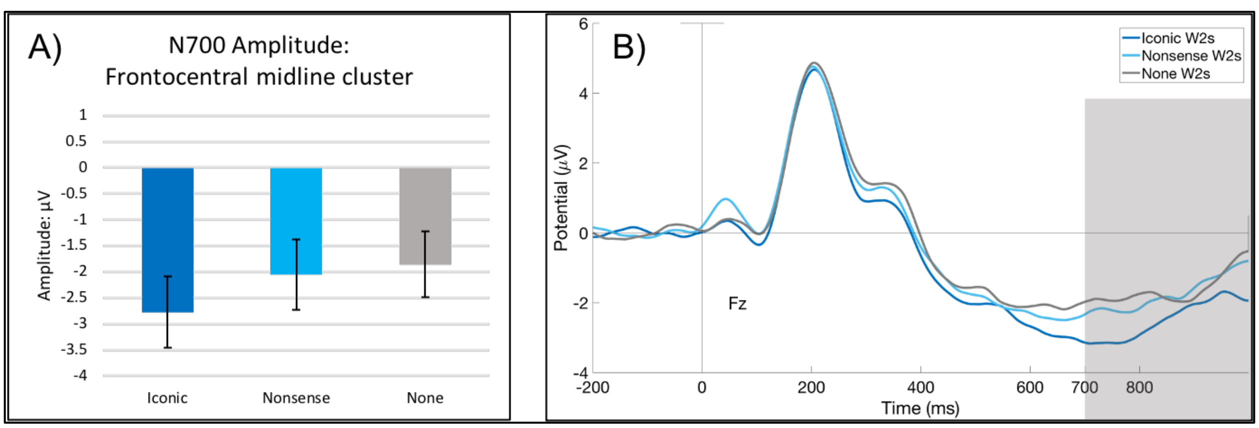


Figure 8. Mean N700 amplitude by gesture condition. The effect of gesture on N700 peak latency at a frontocentral midline cluster (AFz, Fp1, Fp2, Fz, F3, F4, Fc1, Fc2, and Cz) was not significant. A) Mean N700 amplitude values are plotted by condition, which did not significantly differ by gesture type. B) An ERP waveform of the mean amplitude at electrode Fz (selected from the cluster) for all three conditions, with time plotted on the x-axis, voltage plotted on the y-axis, and the measured time window (700-1000 ms) shaded.

**Mass univariate analysis**

Figure 9 depicts the results of our omnibus mass univariate analysis, which included all electrodes and all 100 ms time bins from 0-1000 ms. The cluster-based mass permutation analysis revealed a significant difference between W1 gesture types ($p = 0.01$) over primarily central and parietal electrodes with a temporal extent of 500-800 ms, a temporal mass peak of 700 ms, and a spatial mass peak at electrode CP2. Follow-up pairwise comparisons were conducted examining the effect of gesture on W2 amplitude.

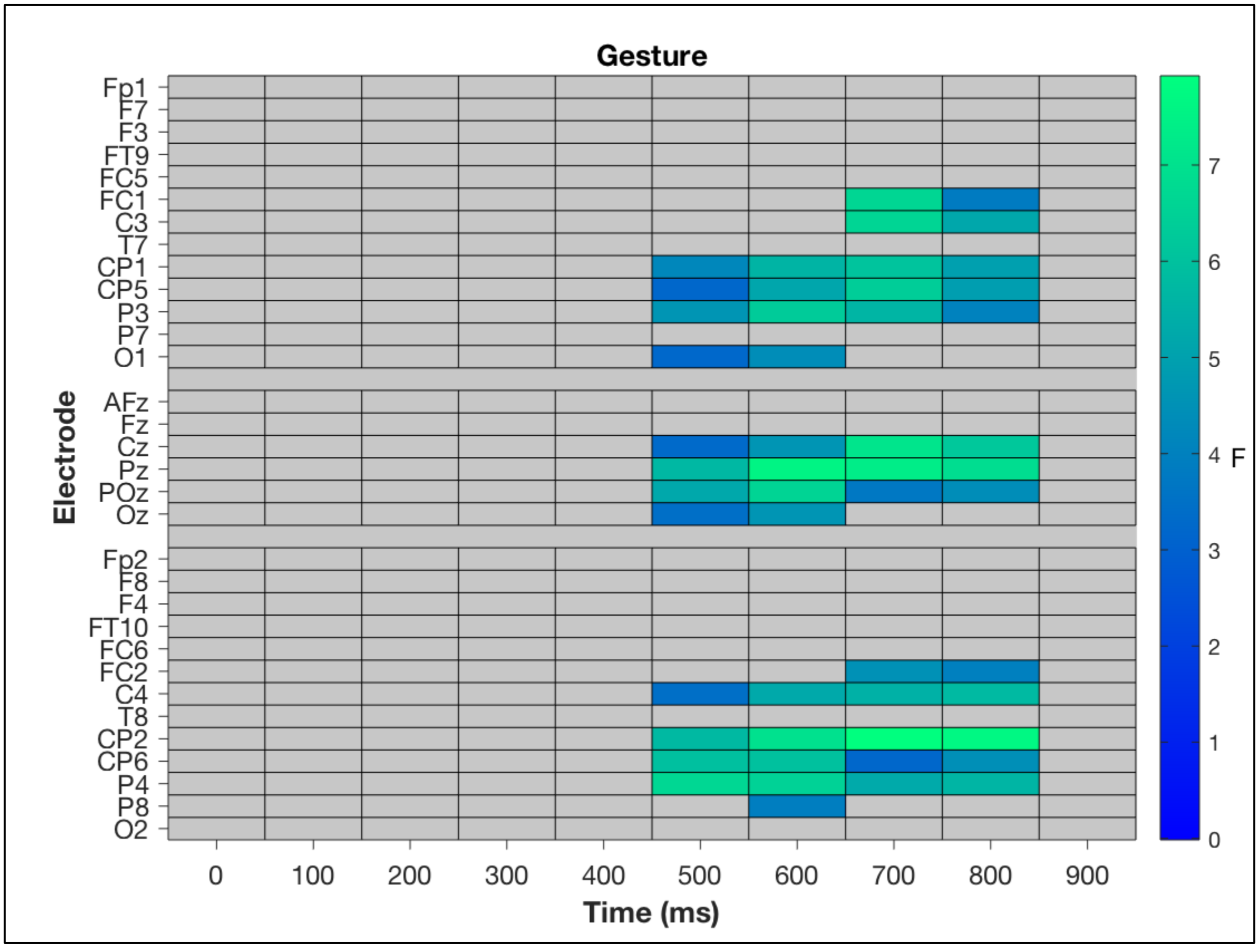


Figure 9. Mass Univariate Analysis results. A mass univariate analysis raster plot revealed a significant difference between W2 gesture types, primarily around central and parietal electrodes with a temporal extent of 500-800 ms, a temporal mass peak of 700 ms, and a spatial mass peak at electrode CP2.

The results of the pairwise comparisons are described in Table 1 and representative electrodes and topographical distributions are plotted in Figure 10. One significant cluster emerged from the iconic vs. no gesture comparison ($p < 0.001$) between 600-800 ms with a temporal peak at 700 ms and a spatial mass peak at electrode CP2. This effect was smaller (less positive) for iconic relative to no gesture trials. Another significant cluster emerged between 600-900 ms in the nonsense vs. no gesture comparison ($p = 0.004$) with a temporal peak at 600 ms and a spatial mass peak at electrode Pz. Amplitude was smaller (less positive) for nonsense relative to no gesture trials. No significant cluster emerged from the comparison of iconic and nonsense gesture conditions.

Table 1. Characterization of significant clusters in the Mass Univariate Analysis.

| Time window (ms) | Effect | Cluster *p*-value | Spatial extent | Temporal extent (ms) | Spatial cluster mass peak | Temporal cluster mass peak |
|---|---|---|---|---|---|---|
| 600 – 900 | Gesture: None > Iconic | 0.0005 | FC1, FC2, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4 | 600 – 800 | CP2 | 700 ms |
| 600 – 900 | Gesture: None > Nonsense | 0.0035 | Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4 | 600 – 900 | Pz | 600 ms |

Table 1. All identified electrodes contained at least one significant time point and all identified time points contained at least one significant electrode site. W2s in the no gesture condition had increased (more negative) amplitudes than W2s in iconic and nonsense gesture conditions. There was no difference in W2 amplitude between iconic and nonsense gesture conditions.

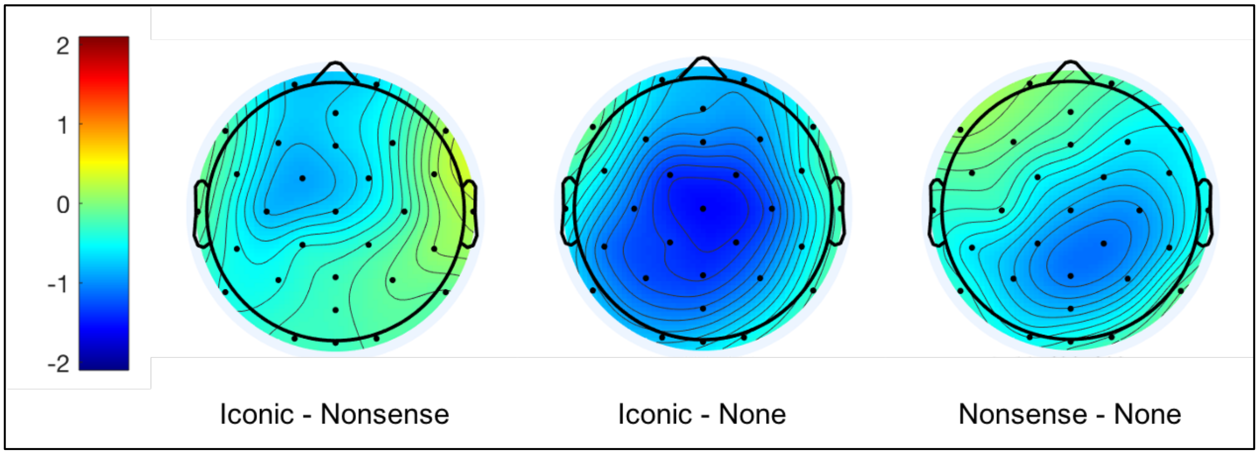

Figure 10. Differences in amplitude between gesture conditions. Topographical scalp maps show the difference in W2 amplitudes between W1 gesture conditions between 600-800 ms. W2 amplitude was greater (more negative) for non-gestured compared to gestured conditions, and there was no difference between iconic and nonsense gesture conditions.

In sum, W2s that had been paired with either iconic-gestured or nonsense-gestured W1s evoked a late-onset (~600 ms) positivity that was largest over centroparietal electrodes. We address this unexpected finding in the discussion.

**Exploratory analyses**

As previously mentioned, visual inspection of the grand averaged waveforms revealed patterns that resembled a frontally distributed N700 effect between 600-800 ms. We conducted an exploratory within-subjects repeated measures ANOVA of the frontocentral midline cluster (AFz, FP1, FP2, Fz, F3, F4, FC1, FC2, and Cz) during this window. The difference between gesture conditions was significant, [$F(2, 58) = 3.06$, $p = 0.05$, $\eta_p^2 = 0.1$]. Consistent with our hypothesis regarding N700 potentials, iconic gestures elicited greater (more negative) amplitudes of W2s than no gestures during this time window [$t(29) = -2.34$, $p = 0.03$, *Cohen's d* = -0.28]. There was no significant difference between iconic and nonsense gestures [$t(29) = -1.5858$, $p = 0.1236$, *Cohen's d* = -0.17] nor nonsense and no gestures [$t(29) = -0.91412$, $p = 0.3682$, *Cohen's d* = -0.1] (Figures 11a and 11b).

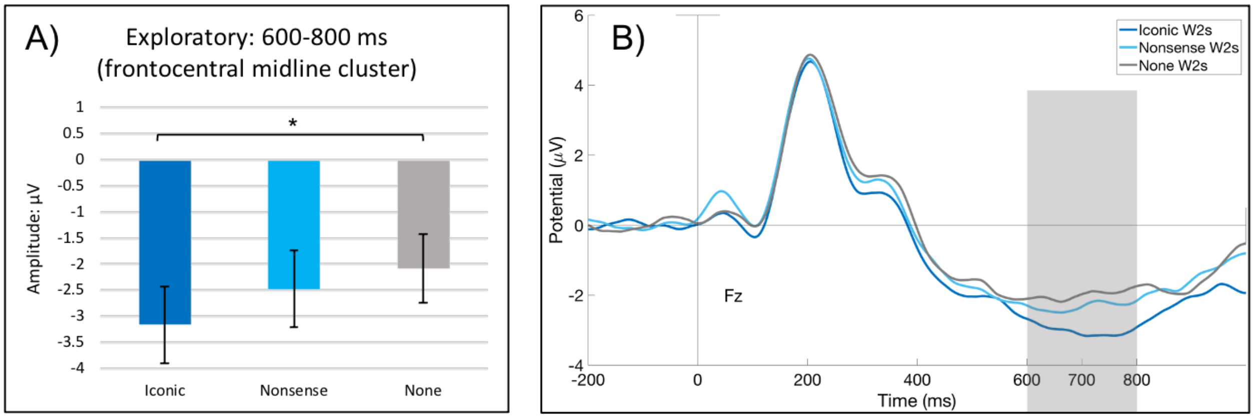


Figure 11. Exploratory ERP analyses. Exploratory analyses of a frontocentral midline electrode cluster between 600-800 ms revealed a significant effect of W1 gesture on W2 mean amplitude. A) W1 iconic gestures elicited increased amplitudes of W2s compared to no gestures. There was no significant difference between iconic and nonsense gestures conditions, nor nonsense and no gesture conditions. B) An ERP waveform of the mean amplitude at electrode Fz (selected from the cluster) for all three conditions, with time plotted on the x-axis, voltage plotted on the y-axis, and the measured time window (600-800 ms) shaded.

## Discussion

Co-speech hand gestures are a naturally occurring part of human communication and facilitate social interaction, learning, and memory. However, it remains unclear as to which cognitive processes are influenced by gesture, and whether different types of gesture are more effective at improving memory than others. In the current study, we recorded continuous EEG data while participants watched videos of an actor reciting sentences, and tested whether co-occurring iconic, nonsense, or no gestures enhanced mental imagery and influenced item and associative memory for novel word pairs heard at the end of each trial.

Analysis of free recall data revealed that iconic gestures improved item memory for the gestured word (W1) compared to nonsense or no gestures. However, these mnemonic benefits did not extend to W2 recall, nor to recall of entire pairs, as would be expected by Dual Coding Theory. In fact, the probability of W2 recall conditioned on W1 recall was significantly *lower* for iconic-gestured pairs compared to non-gestured pairs. This lack of associative memory benefit contradicts our previous work (Cairney et al., in preparation), in which W1 iconic gestures led to improvements in both item and associative memory for word pairs relative to beat- and no-gestured word pairs. The primary difference was the use of beat gestures in our prior study—small hand movements that do not convey meaning but are thought to highlight salient parts of discourse—which were replaced by nonsense gestures in the current experiment. Taken together, this pattern of results suggests that the inclusion of the nonsense gestures caused a more general change in the way participants processed W1 gestures that impacted iconic as well as nonsense gestures. One possibility is that participants diverted more cognitive resources toward gesture processing overall due to presence of ambiguous gestures, thereby leaving fewer resources for

attending to and encoding the corresponding W2s. We will return to this possibility when discussing the electrophysiological results.

Although iconic-gestured word pairs were no more likely to be remembered than nonsense- or no-gestured word pairs, participants did rate iconic-gestured pairs as easier to imagine than word pairs in the other two conditions. Moreover, analyses of the relationship between imagery ratings and memory collapsed across gesture type revealed significantly higher ratings for correctly recalled word pairs when compared to all other recall types (singleton W1s, singleton W2s, forgotten). Taken together, these findings raise the possibility that the data contain an *indirect* effect of gesture on associative memory that is mediated by the tendency for iconic gestures to enhance pair imageability. Although our analysis strategy did not permit us to directly test for this indirect effect, this explanation would be consistent with Dual Coding Theory's emphasis on imagery as a key mechanism behind the creation of "conceptual pegs" that facilitate associative memory. In this case, the lack of a direct effect on W1 gesture type may indicate that another consequence of W1 iconic gestures (such as the increased cognitive load described above) had an opposing effect on associative memory. Future research using multilevel mediation could help to corroborate this account.

Our a priori hypotheses regarding effects of W1 gesture on W2 ERPs were based on prior literature implicating P3 and N700 ERP components in attentional orienting and mental imagery, respectively. Specifically, enhanced P3 effects for either gesture type would support an attentional highlighting account of gesture, whereas greater N700 amplitudes for W2s preceded by iconic-gestures W1s would be consistent with facilitated imagery, which is predicted by Dual Coding Theory. Neither effect was found, rendering the results equivocal. That said, exploratory analyses based on visual inspection identified an abbreviated N700-like effect that was larger

(more negative) for iconic-gestured relative no non-gestured word pairs, but was not affected by nonsense gestures. Although this finding was based on exploratory analyses and should be interpreted with caution, it coheres with the presence of higher imageability ratings for iconic-gestured pairs relative to the other two conditions.

Cluster-based mass permutations of the ERP data revealed an unexpected significant effect of the no gesture condition relative to both gesture conditions, which took the form of an increased positivity for non-gestured W2s over centroparietal electrodes during 600-800 ms. Although we can only speculate as to the functional significance of this effect, it has the spatiotemporal characteristics of a late P3b and/or P600 component, both of which have been associated in previous research with attention toward and evaluation of task-relevant information. The P3b is thought to reflect the process of updating the contents of working memory (e.g., “context updating”), which involves the engagement of attentional resources (Donchin & Coles, 1988; Pontifex, Hillman, & Polich, 2009) In general, P3b effects tend to be reported in studies examining target stimulus processing, with larger amplitudes for novel compared to frequent target stimuli (Polich, 2007).[3] However, this effect has been found to reverse as the task become more difficult (Simon et al., 2016), leading some researchers to conclude that maintaining sustained attention appropriates cognitive resources that would otherwise be devoted to context updating (Watter, Geffen, & Geffen, 2001). Thus, a P3b interpretation of the present results would suggest that the addition of gestures to W1s increased the overall cognitive load of the task, which may divert attention away from updating the contents of working memory to incorporate the W2.  This account would explain the observed

[3] The P3a and P3b are subcomponents of the P300. Our *a priori* analysis strategy was more suitable to finding a P3a effect, which is elicited by stimuli that are not task-relevant (for example, in passive tasks or in situations in which the surprising stimulus does not require a response; Polich, 2007).

decreased W2 amplitudes for gestured compared to non-gestured trials as well as the lower rates of associative memory (when defined using the conditional probability of W1|W2) for the gestured conditions compared to the no gesture condition.

P600 effects tend to be reported in language studies in response to syntactic expectancy violations, which require participants to resolve conflict between expectations of syntactic structure based on early word categorization and the syntax expressed (i.e., the syntactic violation). The P600 may not be limited to syntactic processing however, and a modest body of research suggests that the P600 may index error detection and conflict monitoring more globally (Kolk & Chwilla, 2007; Nunez-Pena & Honrubia-Serrano, 2004). In fact, the spatiotemporal and functional similarities between P3b and P600 components have led some researchers to question whether the two are distinct components (Leckey & Federmeier, 2020), or whether both reflect the same domain-general process of context updating. Similar to our P3b speculation, decreased P600 amplitude for W2s following gestured compared to non-gestured W1s may reflect greater demands on attentional resources at the cost of context updating for new information following the gestured word.

Regardless of whether the P600 is an extension of the P3b, our mass univariate analysis results cohere with the notion that fewer resources were dedicated to context updating following the presentation of W2s during trials that contained gestures. It may be the case that iconic gestures reinforced the encoding of W1s by making them more imageable and thus more memorable, but detracted attention from W2s, which may explain why a boost in memory was not observed for iconic-gestured W2s or word pairs. Put differently, iconic gestures may have contributed to the formation of a "conceptual peg" by enhancing W1 imageability, but in a way that detracted from the associated information intended to "hang". Notably, we did not observe

similar behavioral effects in our prior study: iconic-gestured W2s were recalled more often than beat- and non-gestured trials. This suggests that the tendency for the W1 iconic gestures to divert processing resources from the W2s may have been a direct result of the co-mingling of iconic gestures with nonsense gestures, which were ambiguous by design and thus may have been more attentionally demanding. An important avenue for future research is to test this notion directly by recording ERPs in a study in which only iconic and no gesture trials are included.

A potential limitation to this experiment was that our ERP analyses were necessarily limited to the W2s that followed gestured W1s. Due to sensorimotor confounds between gestured and non-gestured trials, we were unable to meaningfully interpret neural responses to the gestures themselves, nor to the co-gestured W1s. Different cognitive processes may be engaged during observation and encoding of meaningful and nonsense gestures as well as words directly accompanied by gestures, and understanding these differences will also be important to fully account for the ways in which gesture can impact memory, including the memory improvement found in the present study for iconic-gestured W1s. For example, Straube and colleagues (2014) reported different neural regions of activation when participants viewed gestures that were related versus unrelated to the accompanying speech, and both of these patterns were associated with better memory relative to non-gestured speech. To address this limitation, future research could use stimuli in which gestures occur just prior to the onset of the W1, rather than concurrently.

In summary, the results of this study provide tentative support for the utility of Dual Coding Theory as a framework through which to understand the mnemonic benefits of co-speech gesture. More specifically, meaningful gestures mimicked the effects of word concreteness by both improving item memory and enhancing imageability for concurrent speech. Although we

did not find direct evidence that W1 iconic gestures enhance associative memory, both ERP and behavioral analyses suggest that an indirect effect may have been present that was mediated by imageability but not detectable by our analysis methods. By contrast, our results were largely incompatible with an attentional highlight explanation. Not only did the nonsense gestures fail to improve memory, but the reduced P3b-like amplitudes for both gesture conditions raise the possibility that both gesture types diverted cognitive resources away from the W2s. These results contribute to the extant body of co-speech gesture research and may help to inform behavioral interventions with clinical populations who suffer from memory deficiencies or social and communicative problems.

## References


Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. Academic Press.

Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. Brain and language, 125(1), 47-53.

Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture*, *1*(2), 129-149.

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, *124*(2), 143-152.

Cairney, B.E., West, S.H., & Lucas, H.D. (*manuscript in preparation*). Neural mechanisms of speech-gesture integration and downstream effects on memory.

Church, R. B., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture*, *7*(2), 137-158.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2010). Gesturing makes memories that last. *Journal of memory and language*, *63*(4), 465-475.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and clinical neurophysiology*, *39*(2), 131-143.

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.

Dimitrova, D., Chu, M., Wang, L., Özyürek, A., & Hagoort, P. (2016). Beat that word: How listeners integrate beat gesture and focus in multimodal speech discourse. *Journal of cognitive neuroscience*, *28*(9), 1255-1269.

Donchin, E. (1981). Surprise!… surprise?. *Psychophysiology*, *18*(5), 493-513.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating?. *Behavioral and brain sciences*, *11*(3), 357-374.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. *Event-related brain potentials in man*, *349*, 411.

Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, *45*(3), 445-454.

Duncan-Johnson, C. C. (1981). Young Psychophysiologist Award Address, 1980: P300 Latency: A New Metric of Information Processing. *Psychophysiology*, *18*(3), 207-215.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological psychology*, *14*(1-2), 1-52.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in psychophysiology*, *2*(S 1), 78.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Fields, E. C. (2017). Factorial Mass Univariate ERP Toolbox [Computer software]. Available from: https://github.com/ericcfields/FMUT/releases.

Fields, E. C., & Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. Psychophysiology, 57(2), e13468.

Folstein, J. R., & Van Petten, C. (2011). After the P3: late executive processes in stimulus categorization. Psychophysiology, 48(6), 825-841.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in cognitive sciences*, *3*(11), 419-429.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*(12), 1711-1725.

Gullick, M. M., Mitra, P., & Coch, D. (2013). Imagining the truth and the moon: an electrophysiological study of abstract and concrete word processing. *Psychophysiology, 50*(5), 431-440.

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language*, *28*(5), 594-609.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological bulletin*, *137*(2), 297.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of cognitive neuroscience*, *19*(7), 1175-1192.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and language*, *89*(1), 253-260.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and language, 100*(3), 257-261.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. Science, 197, 792–795.

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600 (s): Positive contributions to language comprehension. Psychophysiology, 57(7), e13351.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience, 8*, 213.

Lucas, H. D., Hubbard, R. J., & Federmeier, K. D. (2017). Flexible conceptual combination: Electrophysiological correlates and consequences for associative memory. *Psychophysiology*, *54*(6), 833-847.

Luck, S. J., Fan, S., & Hillyard, S. A. (1993). Attention-related modulation of sensory-evoked brain activity in a visual search task. *Journal of cognitive neuroscience*, *5*(2), 188-195.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology, 54(1), 146-157.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of educational psychology*, *83*(4), 484.

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, *18*(5), 904.

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience*, *28*(47), 12539-12545.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Neville, H., Snyder, E., Woods, D., & Galambos, R. (1982). Recognition and surprise alter the human visual evoked response. *Proceedings of the National Academy of Sciences*, *79*(6), 2121-2123.

Núñez-Peña, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, *18*(2), 130-141.

Özyürek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of cognitive neuroscience*, *19*(4), 605-616.

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 32-38.

Paivio, A. (1969). Mental imagery in associative learning and memory. Psychological review, 76(3), 241.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive psychology*, *5*(2), 176-206.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

Polich, J., & Donchin, E. (1988). P300 and the word frequency effect. *Electroencephalography and clinical neurophysiology*, *70*(1), 33-45.

Pontifex, M. B., Hillman, C. H., & Polich, J. (2009). Age, physical fitness, and attention: P3a and P3b. *Psychophysiology*, *46*(2), 379-387.

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1311.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall?. *Language and Cognitive Processes*, *27*(5), 665-681.

Straube, B., Meyer, L., Green, A., & Kircher, T. (2014). Semantic relation vs. surprise: The differential effects of related and unrelated co-verbal gestures on neural encoding and subsequent recognition. *Brain research*, *1567*, 42-56.

Thompson, L. A., Driscoll, D., & Markson, L. (1998). Memory for visual-spoken language in children and adults. *Journal of Nonverbal Behavior*, *22*(3), 167-187.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998-1003.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: an electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*(6), 1024-1037.

Wu, Y. C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. Psychonomic bulletin & review, 14(1), 57-63.

Yates, F. A. (2013). Art of memory. Routledge.

# Vita

Brianna E. Cairney was born in Seattle, Washington. She received an Associate's of Science degree in 2009 and worked in the professional sector for a number of years. She returned to school and upon graduating from the University of Washington in 2016, accepted a post baccalaureate research position in Dr. Raphael Bernier's Autism research lab. Brianna entered the Cognitive and Brain Sciences doctoral program at Louisiana State University in 2018 and began working with Dr. Heather Lucas in her Brain and Memory Laboratory, researching the mnemonic benefits of co-speech hand gestures. Brianna plans to receive her Masters in December 2021 and will continue to work with Dr. Lucas to complete her Doctorate.