

Published in final edited form as:

Brain Lang. 2011 December ; 119(3): 184–195. doi:10.1016/j.bandl.2011.07.002.

Are depictive gestures like pictures? Commonalities and differences in semantic processing

Ying Choon Wu^{a,b} and Seana Coulson^{a,c}

Ying Choon Wu: ywu@cogsci.ucsd.edu; Seana Coulson: coulson@cogsci.ucsd.edu

^aCenter for Research in Language, UC San Diego 0526, 9500 Gilman Dr., La Jolla, CA 92093

^bSwartz Center for Computational Neuroscience UC San Diego 0559, 9500 Gilman Dr., La Jolla, CA 92093

^cUC San Diego, Dept. of Cognitive Science 0515, 9500 Gilman Dr., La Jolla, CA 92093

Abstract

Conversation is multi-modal, involving both talk and gesture. Does understanding depictive gestures engage processes similar to those recruited in the comprehension of drawings or photographs? Event-related brain potentials (ERPs) were recorded from neurotypical adults as they viewed spontaneously produced depictive gestures preceded by congruent and incongruent contexts. Gestures were presented either dynamically in short, soundless video-clips, or statically as freeze frames extracted from gesture videos. In a separate ERP experiment, the same participants viewed related or unrelated pairs of photographs depicting common real-world objects. Both object photos and gesture stimuli elicited less negative ERPs from 400–600ms post-stimulus when preceded by matching versus mismatching contexts (dN450). Object photos and static gesture stills also elicited less negative ERPs between 300 and 400ms post-stimulus (dN300). Findings demonstrate commonalities between the conceptual integration processes underlying the interpretation of iconic gestures and other types of image-based representations of the visual world.

Keywords

gesture; semantic processing; N400; N300; object recognition; action comprehension; gesture N450

1. Introduction

Depictive, or iconic, gestures are spontaneous body movements that enhance ongoing speech by representing visuo-spatial properties of relevant objects and events. Much of their capacity for signification derives from perceptual, motoric, and analogical mappings that can be drawn between gestures and the conceptual content that they evoke. For example, a speaker might indicate the size of a bowl by holding his hands apart from each other,

© 2011 Elsevier Inc. All rights reserved.

Corresponding Author: Ying Wu, Center for Research in Language, UC San Diego, 9500 Gilman Drive 0526, La Jolla, CA 92093-0526, ywu@cogsci.ucsd.edu, phone (858) 869-7014, FAX: (858) 822-5097.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

creating perceptual similarity between the span of the bowl and the span of open space demarcated by his hands. Alternatively, he might cup his hands as though holding the bowl, showing something about the bowl's size and shape from the depicted act of holding it.

Several theoretical views have either implicitly or explicitly pointed to similarities between the processes engaged during the comprehension of depictive gestures and those mediating the comprehension of pictures or bona fide objects. It has been argued, for example, that depictive gestures are like images in that they afford the opportunity to encode global, holistic relations, which contrast with the analytic, linearly segmentable properties of speech (McNeill, 1992). Along a similar line of reasoning, Kendon (2004) writes, "...descriptive gestures, rather like drawings or pictures, can achieve adequate descriptions with much greater economy of effort and much more rapidly than words alone can manage" (p. 198). Finally, Feyereisen and deLannoy (1991) suggest that gesture comprehension, "...may be compared to other kinds of visual processing like object recognition" (p. 90).

Experimental research on gesture provides indirect support for commonalities between the cognitive and neural basis of gesture comprehension and object recognition. For example, seeing a speaker's gestures benefits performance on certain tasks that require the listener to conceptualize spatial relations, such as drawing abstract designs (Graham & Argyle, 1975) or selecting targets from an array (McNeil, Alibali, & Evans, 2000) on the basis of the speaker's description. One possible interpretation of these outcomes is that individuals formulate mental "pictures" of key task-related visuo-spatial features. Further evidence in favor of this idea was advanced by the finding that depictive co-speech gestures impact the recognition and semantic analysis of pictures presented immediately after a speaker's utterance (Wu & Coulson, 2007a; Wu & Coulson, 2010a, 2010b).

The present study undertakes a more direct investigation of the relation between object recognition and the comprehension of depictive gestures by comparing ERPs elicited during the semantic processing of these two stimulus types. Our hypothesis that understanding gestures and objects involves at least partially overlapping cognitive systems predicts that ERP components previously implicated in object recognition should also be observed in the comprehension of depictive gestures. Moreover, by comparing the scalp distribution, or topography, of ERP responses elicited by objects and gestures, it is possible to draw inferences about the degree of similarity between the neurocognitive systems engaged by each of these representational media.

Notably, this line of investigation diverges from much current research focusing on the relationship between gesture and action. According to the popular mirror theory of action comprehension (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), understanding purposive actions engages some of the same neural populations and brain systems that are recruited during action production, including the inferior frontal gyrus (BA 44/45) and posterior parietal cortex (Buccino, Binkofski, & Riggio, 2004; Iacoboni, 2009). Because gestures resemble overt action in that they are performed with the hands and body, a link has been drawn between mirror systems, gestures, and the evolution of human language (Corballis, 2003; Rizzolatti & Arbib, 1998; Rizzolatti & Craighero, 2004). Recent neuroimaging research has revealed that understanding gestures with concurrent speech does indeed lead to activation of "classic" mirror regions of the brain, such as the inferior frontal gyrus (Holle, Gunter, Ruschemeyer, Henenlotter, & Iacoboni, 2008; Willems, Ozyurek, & Hagoort, 2007, 2009).

In exploring commonalities and differences between object recognition and the interpretation of depictive gestures, the present study seeks to elucidate how meaning is constructed from the highly schematic, transient, and conceptually complex features of

depictive gestures. Whereas more robust and invariant depictions of objects afforded by actual pictures are thought to directly activate stored knowledge by virtue of their similarity to their referent, depictive gestures usually require some form of supportive context – typically concurrent speech – in order for the comprehender to draw mappings between their visuo-spatial properties and their meaning (e.g. to interpret hand shape as indicating holding a bowl or hand position as indicating the size of the bowl). Nevertheless, in spite of the non-obvious significance of a depictive gesture's semiotic elements, it is possible that comprehenders attempt to integrate them in a manner analogous to the features of more robust pictorial representations such as photographs.

Previous electrophysiological research has uncovered two correlates of image-based semantic analysis. Specifically, in pictorial priming paradigms, drawings and photos have been shown to elicit an anterior negativity peaking around 300ms after stimulus onset (N300), as well as a more broadly distributed negativity peaking approximately 400ms post-stimulus (N400) (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999). For both time windows, the difference waves created by subtracting ERPs to contextually congruent pictures from incongruent ones (to be called the dN300 and dN400 hereafter) tend to be negative going, indicating that incongruent stimuli elicit N300 and N400 of greater amplitude than congruent ones.

The dN400 response to incongruent images has been related to the “classic” N400 observed in response to linguistic stimuli. A well-studied ERP component, the lexical N400 is generally thought to index the degree of semantic fit between a word and its preceding context (Wu & Coulson, 2005). Words that are expected given their context typically elicit a reduced N400, whereas unexpected ones elicit a large amplitude N400 (Kutas & Hillyard, 1980, 1984).

Striking parallels between the word and picture dN400 include sensitivity to sentential congruity and global discourse coherence (Federmeier & Kutas, 2001, 2002; Ganis, Kutas, & Sereno, 1996; van Berkum, Hagoort, & Brown, 1999; West & Holcomb, 2002; Willems, Ozyurek, & Hagoort, 2008). Further, just as pseudo-words elicit larger N400s than unrelated words (Holcomb, 1988), unrecognizable images elicit larger N400s than recognizable ones (McPherson & Holcomb, 1999). Germane to the field of gesture studies, it has also been shown that both related words (Kelly, Creigh, & Bartolotti, 2009; Kelly, Ward, Creigh, & Bartolotti, 2007; Lim et al., 2009; Wu & Coulson, 2007b) and pictures (Wu & Coulson, 2007a) primed by iconic gestures elicit N400 of decreased amplitude relative to unrelated ones.

Given these similarities in time course and functional characterization, the word and picture dN400 have been proposed to reflect the activity of similarly functioning neural systems mediating the comprehension of language and images, respectively (Ganis et al., 1996; Willems et al., 2008). However, the dN400 effect elicited by pictures tended to be largest over the front of the head, whereas the dN400 produced by words tended to be largest at centro-parietal sites (Ganis et al., 1996), as would be expected if overlapping, but non-identical, neural systems are responsible for picture and word comprehension.

Additionally, an earlier negative-going waveform that peaks around 300 ms after stimulus onset (N300) in response to images also suggests that some processes are specific to image comprehension¹. Like the N400, the N300 is modulated by contextual congruity, but is not

¹Some researchers have reported an N300-like effect in response to words. Cristescu and Nobre (2008), for instance, report a broadly distributed N300 component elicited by words denoting concrete objects, such as tools and animals. This finding is consistent with the view that lexical and visual representations of concrete objects engage at least partially overlapping cognitive systems.

sensitive to gradations of relatedness (Hamm, Johnson, & Kirk, 2002; McPherson & Holcomb, 1999). Further, whereas the dN300 is largest over anterior electrode sites, the dN400 to pictures tends to be more broadly distributed (Hamm et al., 2002; McPherson & Holcomb, 1999).

On the basis of these functional and topographic differences, a number of researchers have proposed that the picture N300 and N400 index different aspects of the semantic processing of images. Whereas the N300 family of potentials has been proposed to reflect the process of matching perceptual input with image-based representations in long-term memory (Schendan & Kutas, 2002; Schendan & Kutas, 2007; Schendan & Maher, 2009; West & Holcomb, 2002), the N400 family of potentials has been related to the integration of semantic information across a broad range of input modalities.

While image comprehension has been extensively studied with ERPs over the past decades, depictive gestures have only recently begun to receive attention from cognitive neuroscience (Kelly, Manning, & Rodak, 2008; Kelly & Ozyurek, 2007). Wu & Coulson (2005) recorded ERPs time-locked to the onset of depictive gestures produced during spontaneous discourse in which a speaker described cartoons. In this experiment, all gestures were preceded by excerpts of the original cartoons presented to the speaker. On congruent trials, gestures were paired with the actual cartoon segment that the speaker had been describing. On incongruent trials, gestures were preceded by cartoon segments that bore no relation to the topic being described by the speaker.

It was found that both congruent and incongruent gestures elicited a negative-going deflection of the waveform peaking around 450ms after stimulus onset (gesture N450), with enhanced negativity for incongruent items. Because of their similarities in time course and sensitivity to contextual congruity, the gesture N450 was construed as a member of the N400 class of negativities, and proposed to index the integration of gesturally-based semantic information with preceding context. No N300 effects were observed.

The absence of a dN300 effect may be a result of the visual complexity of dynamic gestures as compared to static representations of objects. It is possible that the processes indexed by the dN300 are delayed in onset during gesture comprehension, overlapping with those indexed by the dN450. If object recognition processes engaged by pictures and photographs are also sensitive to gestural representations, visually simpler “static gesture snapshots” may yield earlier, more discernible N300-like activity.

To explore this idea, the present study used static gesture freeze frames extracted from dynamic gesture video clips. Although static representations of depictive gestures contain considerably less information than dynamic ones, jpegs were selected that preserved important semiotic cues, such as hand shape, body configuration, and hand location. Thus, understanding these gesture “snapshots” was likely to involve similar visual analysis and integration processes mediating the understanding of full gestures. Trials were constructed by pairing cartoon contexts with congruent and incongruent dynamic and static gestures, with each gesture following the presentation of a cartoon. Participants viewed a cartoon-gesture pair, and then judged whether a subsequent probe word was related to either of the preceding stimuli (see Figure 1A). ERPs were time locked to the onset of the gestures.

Because this study focuses on the interpretation of gestural forms rather than speech-gesture integration, it was deemed preferable to use the original cartoon clips rather than the speaker’s verbal descriptions of them as contextual support. In this way, a communicative situation was approximated in which both speaker and listener shared a common referential framework.

To compare ERP effects elicited by static gestures with those elicited by more conventionally meaningful visual representations, a second experiment was conducted in which the same participants viewed related or unrelated pairs of photographs depicting common household objects from the stimulus corpus used in McPherson and Holcomb (1999). In keeping with Experiment 2 of their original study, probe images were either related or unrelated to their preceding prime, or were unidentifiable. In the present study, participants viewed a series of two photographs followed by a probe word (see Figure 1B). Participants judged whether the word was related to either of the preceding photographs, and ERPs were time locked to the onset of the second photograph.

The similarity of ERP congruity effects for static and dynamic gestures versus photographs of objects afforded a basis for comparing cognitive systems mediating gesture and object comprehension. Because both of these stimulus types were presented in the visuo-spatial modality, one might expect N300 and N450 effects time-locked to single gesture jpegs to exhibit an anterior focus, analogous to that observed in response to object photographs and other types of representations thought to engage image specific processes, such as picture stories (West & Holcomb, 2002), videos (Sitnikova, Kuperberg, & Holcomb, 2003), and concrete words (Holcomb, Kounios, Anderson, & West, 1999; Kounios & Holcomb, 1994). On the other hand, because gestures are schematic symbols used primarily in conjunction with language, congruity effects elicited by static gestures could prove more prominent over the back of the head, in keeping with the idea that gesture meanings were processed in a manner analogous to words. Such an outcome would be consistent with results described in Ozyurek et al (2007), who reported a similar time course and topographical distribution for N400 effects elicited either by incongruent words or gestures in a sentence context.

2. Methods

2.1 Participants

Sixteen volunteers were paid \$24 or awarded course credit for participation. All were healthy, right-handed, fluent English speakers without history of neurological impairment. Their mean laterality quotient, which is derived from the Edinburgh Inventory (Oldfield, 1971) and provides an index of handedness preference, was .7 (SD = .13) (with maximal right handedness indicated by a score of 1).

2.2 Materials

Stimuli were constructed from the same corpus of depictive co-speech gestures used in an earlier study in our laboratory (Wu & Coulson, 2005). An individual was videotaped while describing segments of popular cartoons (Bugs Bunny, Road Runner, and so forth). He was told that the experimenters were creating stimuli for a memory experiment and was unaware of the intent to elicit spontaneous gestures. Instances in which his utterance was accompanied by depictive gestures were digitized into short, soundless video clips. To create static gestures, a single frame (in the form of a jpeg file) was isolated from the video stream of each gesture. Typically, freeze frames were extracted just at the onset of the meaningful phase of movement – that is, the stroke (Kendon, 1972) phase of the gesture – or during a pause between strokes, in the case of iterative or complex gestures, in order to avoid blurring due to rapid motion of the hands.

160 cartoon clips were paired either with dynamic gesture video clips or static gesture freeze frames. Congruent trials were those in which a gesture was paired with the original cartoon clip that it was intended to describe. Incongruent trials involved mismatched cartoon-gesture pairings. Each cartoon-gesture pair was followed by a task word, which participants were asked to judge either as related or unrelated to any preceding element on that trial (on trials

designed to elicit a YES response, the task word was always related at least to the cartoon). This procedure provided a way to assess participant comprehension without requiring explicit decisions during the presentation of the stimulus of interest.

A normative study was conducted to ensure that congruent and incongruent static trials were reliably distinguished as such. (See Wu and Coulson (2005) for details on normative data collected for dynamic gestures). Six volunteers received academic credit for rating the degree of congruity between static gestures and cartoons on a five point scale. On average, the congruity rating was 3.3 (SD = 1) for congruent trials and 1.9 (SD = 0.4) for incongruent ones. A two-tailed matched pairs *t* test revealed that this difference was statistically reliable, $t(159) = 14.5$, $p < .001$.

Eight lists of gesture stimuli were constructed, each containing 80 congruent and 80 incongruent trials (40 static and 40 dynamic of each type), and 80 related and 80 unrelated words. Trials were divided into 8 blocks, each containing a randomized selection of static and dynamic items presented in an interleaved fashion. No cartoon, gesture, or word was repeated on any list, but across lists, each gesture appeared once as a congruent stimulus and once as an incongruent one. Words also appeared once as a related item and once as an unrelated one following all four types of cartoon-gesture pairs (congruent dynamic, congruent static, incongruent dynamic, incongruent static).

For the experiment involving photographs of common objects, 90 related and 90 unrelated image pairs were constructed from digital photographs. Related images shared semantic mappings along a variety of dimensions, including category membership (PANTS-SHIRT), associativity (BACON-EGGS), function (WATER COOLER-CUP), and part to whole (TREE-LEAF). Thirteen volunteers received academic credit for rating the degree of relatedness between each pair of pictures on a five point scale. Ratings for related targets (mean = 4.5; SD = .21) were reliably higher than those for unrelated ones (mean = 1.5; SD = .25) ($t(12) = 27$, $p < .0001$).

Ninety additional trials were constructed by pairing photographs of common objects with unidentifiable targets. The majority of images depicted a single object against a neutral background, though a few objects were photographed in their natural environment (e.g., a tree). As in the gesture stimuli described above, a related or unrelated probe word followed each image pair.

Six lists of object photograph stimuli were constructed, containing 30 related, 30 unrelated, and 30 unidentifiable trials combined into three randomized blocks. No picture prime or probe was repeated on any list. Across lists, however, each identifiable probe was paired both with a related and an unrelated image prime. Further, across lists, each picture prime was paired once with an unidentifiable probe. For all three types of trials, task words were counterbalanced such that each trial was followed once by a related and once by an unrelated word.

2.3 Procedure

Volunteers participated in the two experiments consecutively, over the course of one recording session. Gesture trials began with a fixation cross, presented in the center of a 17 in. color monitor for two seconds. The cartoon and gesture clips were presented at a rate of 48ms per frame with a 600ms pause before the onset of the gesture (in order to allow participants time to establish central fixation). Although cartoons varied in length (mean = 2949ms, SD = 900ms), the duration of each gesture was exactly 2300 ms. One second after the offset of the gesture, a probe word either related or unrelated to the preceding context was presented (see Figure 1A). A short pause (approximately 5–6s) followed each trial as

the next set of video frames was accessed by the presentation software. All video frames were centered on a black background and subtended approximately 10° visual angle horizontally and 7° vertically.

Participants were told that they would watch a series of cartoon segments, each followed first by a video clip of a man describing either the immediately preceding cartoon, or a different one, and then by a probe word. They were asked to press YES or NO on a button box as soon as they felt confident that the task word matched or did not match the preceding context. Response hand was counterbalanced across subjects. Four additional trials were used in an initial practice block. Participants were presented with all eight blocks of gesture stimuli, followed by a short break, and then the final three blocks of object photos.

Like cartoon-gesture pairs, trials involving photographs of common objects also began with a 2s fixation cross. The prime image was presented for one second, followed by a 250ms interval during which the screen was blank, in keeping with presentation parameters used in Experiment 2 of McPherson and Holcomb (1999). Subsequently, the probe image appeared on the screen for 2.3 seconds, matching the duration of static and dynamic gestures. After 500ms, a fixation cross appeared for 500ms, followed by the task word, which remained on the screen for one second (see Figure 1B). All items were presented in the center of the computer monitor. As before, volunteers were instructed to attend to image pairs, and to press YES or NO on the button box depending on whether or not the word agreed with any element of the preceding context.

2.4 EEG Recording and Preprocessing

Electroencephalogram (EEG) was recorded with tin electrodes at 29 standard International 10–20 sites (Nuwer et al., 1999) (midline: FPz, Fz, FCz, Cz, CPz, Pz, Oz; medial: FP1, F3, FC3, C3, CP3, P3, O1, FP2, F4, FC4, C4, CP4, P4, O2; and lateral: F7, FT7, TP7, T5, F8, FT8, TP8, T6). All electrodes were referenced online to the left mastoid, and impedances maintained below 5 k Ω . Recordings were also made at the right mastoid, and EEG was re-referenced to the average of the right and left mastoids off-line. Ocular artifacts were monitored with electrodes placed below the right eye and at the outer canthi. EEG was amplified with an SA Instrumentation isolated bioelectric amplifier (band pass filtered, 0.01 to 40 Hz) and digitized on-line at 250 Hz.

Artifact-free ERP averages time locked to the onset of gestures were constructed from 100ms before stimulus onset to 920ms after. Trials accurately categorized by participants were sorted and averaged. Critical gesture bins contained 30 trials (± 4) on average. The mean artifact rejection rate was 22% (12% SD) for dynamic gestures and 23% (10% SD) for static ones. Critical picture bins contained on average 25 trials (± 4). The mean artifact rejection rate was 17% (13% SD) in response to identifiable picture probes, and 15% (13% SD) in response to unidentifiable items. The higher rate of artifact contamination in the case of gestures relative to object photos is likely due to increased ocular movements produced as dynamic gestures were viewed.

3. Results

3.1 Word Classification Accuracy

Average d' for classification of related and unrelated words was significantly greater than zero (where zero indicates no detection of signal) for all three categories of stimuli – full gestures ($d' = 2.8$; $t(15) = 14$, $p < .0001$); static gesture freeze frames ($d' = 2.8$; $t(15) = 10$, $p < .0001$); and object photos ($d' = 2.5$; $t(15) = 12$, $p < .0001$).

3.2 ERPs

Figure 2 shows ERPs recorded over midline and medial electrode sites, time-locked to the onset of static and dynamic gestures. For all trials, a large, negative-going onset potential can be observed, peaking around 240ms after stimulus onset, followed by a second negative-going deflection of the waveform peaking around 450ms post-stimulus. In the case of dynamic gestures, effects of gesture congruity can be observed from around 400ms to the end of the epoch (900ms), with enhanced negativity for incongruent items relative to congruent ones. In the case of static gestures, congruity effects begin earlier – around 230ms after stimulus onset. Again, incongruent items continue to elicit more negative ERPs until the end of the epoch.

Figure 3 depicts ERPs elicited by the second item in each of our three types of image pairs in the experiment involving photographs of common objects. The negative-going onset potential visible in all conditions peaks around 100ms after stimulus onset, and is smaller than that observed in response to gestures. Effects of image relatedness are visible by 100ms post-stimulus and continue to the end of the epoch, with unrelated image pairs eliciting more negative ERPs than related ones. Unidentifiable trials are differentiated from unrelated ones between approximately 200 and 600ms after stimulus onset, with unidentifiable items resulting in more negative ERPs than unrelated items.

Gesture congruity and picture relatedness effects were both assessed by measuring the mean amplitude and peak latencies of ERPs time-locked to gesture onset from 300–400ms (N300), 400–600ms (Gesture N450), and 600–900ms – in keeping with intervals utilized in Wu & Coulson (2005). (See online supplemental material for a description of results obtained within the 600–900ms time window). Measurements were subjected to repeated-measures ANOVA, using within-subject factors of Gesture Type (Static, Dynamic), Congruity (Congruent, Incongruent), and Electrode Site (29 levels) for gesture trials, and within-subject factors of Target Type (Related, Unrelated, Unidentifiable) and Electrode Site for picture measurements. For all analyses, original degrees of freedom are reported; however, where appropriate, p-values were subjected to Geisser-Greenhouse correction (Geisser & Greenhouse, 1959).

3.2.1 N300 Time Window

Experiment 1: For gesture trials, an omnibus ANOVA analyzing the mean amplitude of ERPs between 300 and 400ms post-stimulus did not yield main effects of Gesture Type or Gesture Congruity, $F's < 2$, n.s. However, an interaction between these two factors was observed, $F(1,15) = 5.7$, $p < .05$, prompting follow-up comparisons of congruity effects individually within each type of gesture. For static gestures, incongruent trials consistently resulted in more negative ERPs than congruent ones, $F(1,15) = 6$, $p < .05$. However, for dynamic trials, gesture congruity did not reliably modulate ERP amplitudes ($F = .16$, n.s.) during this epoch. This pattern of outcomes is consistent with Figure 2, which shows a distinct N300 effect elicited by static gestures, but no N300 differences between congruent and incongruent dynamic gestures.

Experiment 2: To assess how image relatedness modulated the N300, the mean amplitudes of ERPs elicited by related, unrelated, and unidentifiable pictures were compared. A main effect of Target Type, $F(2, 30) = 12.0$, $p < .05$, $\epsilon = .9$, along with a Target Type \times Electrode Site interaction, $F(56, 840) = 11.5$, $p < .05$, $\epsilon = .10$, indicated a differential N300 response to these three kinds of stimuli. To characterize this effect, pre-planned simple contrasts were performed between related and unrelated picture pairs, and between unrelated and unidentifiable pairs.

Contrasting related and unrelated image pairs demonstrated, as expected, that unrelated trials consistently resulted in larger N300 than related trials (Target Type main effect, $F(1, 15) = 21.3$, $p < .05$; Target Type \times Electrode Site interaction, $F(28,420) = 4.2$, $p < .05$, $\epsilon = .10$). Post hoc analysis of ERPs recorded over medial electrodes suggested this effect was larger over anterior scalp electrodes (Target Type \times Posteriority $F(6,90) = 4.0$, $p < 0.05$, $\epsilon = .27$; see Figure 6A). Contrasting unrelated and unidentifiable trials revealed that unidentifiable images resulted in a larger amplitude N300 than unrelated images (Target Type \times Electrode Site interaction, $F(28,420) = 9.0$, $p < .05$, $\epsilon = .14$). This effect was maximal over Fz and FCz, and was most visible over anterior midline and medial electrode sites. Accordingly, post hoc analysis at medial sites revealed an interaction between Target Type and Posteriority, $F(6,90) = 12.5$, $p < 0.05$, $\epsilon = .27$.

3.2.2 N450 Time Window

Experiment 1: Analysis of ERPs to gestures measured 400–600ms post-stimulus revealed a main effect of Gesture Congruity – collapsed across both types of gesture, $F(1,15) = 12.5$, $p < .05$. This effect occurred due to consistently larger gesture N450 responses to incongruent relative to congruent trials. The main effect of Gesture Congruity was qualified by an interaction with Electrode Site, $F(28,420) = 3.0$, $p < .05$, $\epsilon = .10$. Post hoc analysis of ERPs recorded from medial electrodes suggested this interaction was due in part to the greater magnitude of the congruity effect over right than left hemisphere sites (Gesture Congruency \times Hemisphere: Medial, $F[1,15] = 3.5$, $p = .08$; see Figure 5).

In addition to effects of Gesture Congruity, a main effect of Gesture Type also proved reliable, $F(1,15) = 8$, $p < .05$, as well as an interaction with Electrode Site $F(28, 420) = 3.8$, $p < .05$, $\epsilon = .12$. These effects were driven by the more negative mean amplitude of ERPs elicited by static gestures relative to dynamic ones, irrespective of congruity, over central midline and medial electrodes. Post hoc analysis of ERPs from medial electrode sites revealed a reliable interaction between Gesture Type and Posteriority ($F(6,90) = 5.0$, $p < 0.05$, $\epsilon = .33$).

Experiment 2: An analysis of ERPs to object photos obtained within the same time window revealed a main effect of Target Type, $F(2,30) = 20.6$, $p < .05$, $\epsilon = .70$, and an interaction with Electrode Site, $F(56,840) = 7.0$, $p < .05$, $\epsilon = .12$. A simple contrast between related and unrelated items indicated that unrelated items elicited larger N400 than related items (Target Type main effect, $F(1, 15) = 66.6$, $p < .05$; Target Type \times Electrode Site interaction, $F(28,420) = 6.0$, $p < .05$, $\epsilon = .13$). Post hoc analysis of ERPs recorded from medial sites suggested this effect was frontally focused (Target Type \times Posteriority $F(6,90) = 6.2$, $p < 0.05$, $\epsilon = .35$), and larger over the right than the left hemisphere (Target Type \times Hemisphere $F(1,15) = 5.0$, $p < 0.05$). The contrast between unrelated and unidentifiable images established that unidentifiable items resulted in larger N400 than did unrelated images – again chiefly over anterior electrodes (Target Type \times Electrode Site interaction, $F(28,420) = 3.5$, $p < .05$, $\epsilon = .12$; post hoc analysis of medial sites: Target Type \times Posteriority $F(6,90) = 4.2$, $p < 0.05$, $\epsilon = .32$).

3.3 Comparison between ERP effects on static gestures (Experiment 1) and object photos (Experiment 2)

To compare the magnitude of ERP effects in response to static gestures and common objects, a difference wave was constructed by performing a point by point subtraction of the averaged ERP waveform elicited by congruent trials from that elicited by incongruent trials for data collected at each electrode site. A difference wave of the object photo relatedness effect was computed in similar fashion. Using the same time intervals to assess the N300 and N400, the mean amplitudes of static gesture and common object difference waves were

subjected to repeated-measures ANOVA with the within-subject factors of Stimulus Type (Static Gesture, Common Object) and Electrode Site.

Within the N300 time window (300 to 400 ms post-stimulus), a trend toward a main effect of Stimulus Type was observed, $F(1, 15) = 3.9$, $p = .07$. Subsequently, between 400 and 600 ms post-stimulus, a reliable main effect of Stimulus Type was obtained, $F(1, 15) = 15.4$, $p < .05$). During both the N300 and the N450 measurement intervals, the difference wave of ERPs to object photos was more negative than that elicited by static gestures. Because difference waves were derived by subtracting responses to congruent items from responses to incongruent ones, the more negative amplitude of the object photo difference wave indicates that relative to static gestures, common objects yielded a larger dN400 effect, and possibly a larger, more right lateralized dN300 effect.

4. Topography of static gesture N300 and N400 effects

To assess scalp topography of gesture congruency effects, difference waves of the static gesture N300 and N450 effects were computed and submitted to a repeated-measures ANOVA with the within-subject factors of Measurement Interval (300–400ms, 400–600ms) and Electrode Site. Neither a main effect of Measurement Interval, nor a Measurement Interval \times Electrode Site interaction (F 's < 1) proved reliable, suggesting the static gesture dN300 and dN450 were of similar magnitude, and were distributed in comparable ways across the scalp.

5. ERPs to Probe Words²

Experiment 1—To evaluate the impact of different sorts of gesture primes on the processing of language stimuli, ERPs were time-locked to the probe words following each cartoon-gesture pair. Mean amplitude measurements of ERPs elicited between 300 and 500ms after word onset (a standard measurement window for the classic N400 component) were analyzed with repeated measures ANOVA using factors Gesture Type (Dynamic/Static), Congruity (Congruent/Incongruent), Word Probe Relatedness (Related/Unrelated), and Electrode Site (29 levels). Analysis revealed main effects of Gesture Congruity ($F(1,15)=10.27$, $p<0.01$), Probe Word Relatedness ($F(1,15)=28.89$, $p<0.001$), and an interaction between these factors (Gesture Congruity \times Probe Word Relatedness $F(1,15)=10.28$, $p<0.01$; Gesture Congruity \times Probe Word Relatedness \times Electrode Site $F(28,420) = 3.76$, $p<0.01$, $\epsilon = .13$). Gesture Type – that is, whether the gesture prime was dynamic or static – did not modulate ERPs to probe words ($F(1,15)<1$), and did not interact with any of the other factors (all F s < 1.65).

The interaction between Gesture Congruity and Probe Word Relatedness was followed up with separate post hoc analyses of ERPs to words following congruent and incongruent gestures, respectively. Analysis of words following incongruent gestures revealed no reliable effects. By contrast, analysis of probe words following congruent gestures suggested the related probes elicited significantly less negative ERPs than did unrelated (Probe Word Relatedness $F(1,15) = 49.77$, $p<0.0001$; Probe Word Relatedness \times Electrode Site $F(28,420) = 11.09$, $p<0.0001$, $\epsilon = .14$). As in the overall analysis, Gesture Type did not significantly modulate ERPs to probe words ($F(1,15) = 4.05$, n.s.), and did not interact with any of the other factors (all F s < 1.59). In sum, the contextual congruity of preceding gestures affected

²ERPs were also time locked to the onset of probe words following image pairs in Experiment 2. N400 amplitude was assessed by measuring the mean amplitude of ERPs between 300 and 500ms post-onset and analyzing with repeated measures ANOVA with factors Photo Primes (Related/Unrelated/NoID), Word Probe Relatedness (Related/Unrelated), and Electrode Site (29 levels). This analysis revealed only a robust effect of Word Probe Relatedness ($F(1,14)=14.14$, $p<0.01$), as words that were related to one or more of the preceding object pictures elicited ERPs that were 1.5 microvolts less negative 300–500ms post-word onset than did unrelated words. Reduced amplitude N400 for related relative to unrelated words is in keeping with the classic N400 priming effect.

the N400 probe word relatedness effect, as relatedness effects were present when preceded by congruent, but not incongruent gestures. The dynamicity of gestures, however, did not: following congruent dynamic gestures, related words elicited N400 components 3.5 microvolts less negative in amplitude than did unrelated words; following congruent gesture stills, the relatedness effect was 3.1 microvolts.

4. Discussion

To investigate the real time processing of gestures, we recorded ERPs time-locked to the onset of spontaneously produced depictive gestures preceded by congruent or incongruent contexts. In addition to congruity, the dynamicity of gestures was also manipulated: dynamic gestures consisted of short video clips, whereas static gestures were composed of single still images extracted from each dynamic gesture stream. Relative to congruent items, incongruent static gestures elicited an enhanced, frontally focused negativity peaking between 300 and 400ms post-stimulus (dN300). Between 400 and 600ms post-stimulus, all gestures elicited a negative deflection of the ERP waveform with incongruent items yielding enhanced negativity relative to congruent ones (dN450). Both types of gestures also elicited a late congruity effect (600–900ms), which was much larger, more robust, and more broadly distributed for static items.

In addition to gestures, ERPs were recorded to related and unrelated pairs of common object photos. As expected, unrelated targets resulted in larger N300 and N400 components than related ones, and unidentifiable targets yielded larger N300s and N400s than unrelated ones. Relatedness effects were somewhat more right lateralized than reported in the original study using these stimuli (McPherson & Holcomb, 1999), possibly due to differences in the task that participants were asked to perform – explicit judgments of semantic relatedness in response to the target (McPherson & Holcomb, 1999), versus passive viewing of the target in the case of the present study.

Contextual congruity thus reduced the amplitude of the N300 and N400 components of ERPs elicited both by photographs of objects and photographs of a speaker's depictive gestures. The time course of these two congruity effects was highly similar, beginning by 300ms after stimulus onset and lasting 700ms. Further, the topography of the two effects was somewhat similar, with an anterior focus in both cases, suggesting commonalities in object recognition and gesture comprehension. On the other hand, there was evidence for subtle distributional differences between the two, indicating differences in their underlying cortical substrates.

4.1 Gesture ERP effects

The gesture dN450 effect described here resembles that found by Wu & Coulson (2005) in a previous study using this paradigm (Wu & Coulson, 2005). This component is hypothesized to index semantic integration processes similar to those indexed by the N400 observed in response to the semantic analysis of pictures, and analogous to those underlying the classic N400 elicited by verbal stimuli. Like the picture N400, the gesture N450 was broadly distributed, with largest effects over frontal electrode sites, and was not visible at occipital sites (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999; Mudrik, Lamy, & Deouell, 2010; West & Holcomb, 2002). Further, the time course of the gesture dN450 in the present study was in keeping with that of N400 effects elicited by visually complex scenes (West & Holcomb, 2002), and videos (Sitnikova et al., 2003).

The present study is, to our knowledge, the first to demonstrate an N300 congruity effect for depictive gestures (c.f. work on emblems by Gunter & Bach, 2004). The N300 is a family of frontally distributed negativities elicited by a wide range of static image types, including line

drawings (Barrett & Rugg, 1990; Federmeier & Kutas, 2001, 2002; Hamm et al., 2002), photographs of common objects (McPherson & Holcomb, 1999), faces (Paulmann, Jessen, & Kotz, 2009), conventionally meaningful hand shapes, such as “thumbs up” (Gunter & Bach, 2004), and complex scenes (Mudrik et al., 2010; West & Holcomb, 2002). The amplitude of this component is modulated by contextual congruity, with unrelated items eliciting larger N300s than related ones. It is also modulated by the accessibility of an image, with nonsense objects eliciting larger N300s than identifiable ones (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999).

Also associated with the N300 family of components is a frontal negativity peaking around 350 ms (N350), exhibiting a larger amplitude in response to objects depicted from unusual relative to canonical views (Schendan & Kutas, 2003), and in response to successfully identified picture fragments relative to unidentifiable ones (Schendan & Kutas, 2002). The N350 is responsive primarily to global shapes of objects irrespective of local feature differences, and has been interpreted as indexing object model selection, whereby a perceived image is compared with possible structural representations stored in long term memory (Schendan & Kutas, 2007).

In the present study, the finding of a reduced N300 response to contextually congruent static gestures suggests that corresponding stored structural representations were easier to access in the case of congruent relative to incongruent items. Presumably, cartoon contexts activated certain kinds of object representations, which served to facilitate the processing of subsequent gesture stills depicting hand configurations and body postures that could be mapped to these same representations. If congruent gestures did not affect object recognition processes, the brain would be expected to respond to both congruent and incongruent stills as essentially equally unrelated images of a man, and no dN300 would be observed. dN300 congruity effects in the present study thus suggest the neuro-cognitive system that mediates object recognition is also sensitive to the semiotic properties of depictive gestures.

This is quite remarkable given that similarity mappings between cartoons and gestures derive not from basic featural correspondences, but rather from shared relational structure. As an illustration, consider a cartoon segment in which Nibbles, Jerry’s mischievous young cousin from *Tom and Jerry*, jumps onto the rim of a candlestick and begins chomping at the base of the candle, causing it to topple onto Jerry’s head, much in the manner of a tree being felled (Figure 4). As can be seen, the speaker’s left forearm and extended hand appear to reenact the long, straight shape of the candle, as well as its horizontal orientation. Further, the parallel configuration of his left forearm above his right one is analogous to the parallel relationship between the fallen candle and the plate of doughnuts beneath. Notably, however, these mappings are motivated by *shared relations* either between sets of features, such as the shape and orientation of the speaker’s left forearm and the candle, or between distinct items, such as the speaker’s right and left forearms, and the candle and the table. As visual inspection of Figure 4 will confirm, there are few similarities between the cartoon and the gesture from a purely perceptual perspective. The spatial extent subtended by the candle is considerably greater than that subtended by the speaker’s left forearm and hand; the candle is cylindrical in shape, whereas the speaker’s hand is flat; the candle is red (in the original cartoon), whereas the speaker’s arm is covered by the sleeve of a plaid shirt.

The idea that image-based processes mediate the understanding of iconic gestures may seem to contradict existing research proposing a common underlying system for the semantic integration of information from both language and gesture with preceding context (Ozyurek et al., 2007). However, the present study makes no attempt to claim that image-based processes are recruited to the exclusion of language-based or general semantic ones. Rather, it demonstrates that the integration of abstract relational structure, such as that engendered in

congruency mappings between iconic gestures and their referents, involves neuro-cognitive systems similar to those implicated in the comprehension of other types of visually represented congruency mappings between primes and targets, such as cause and effect (e.g. a mousetrap and a dead mouse), function (e.g. a hammer and nail), shared category membership (e. g. pants and a shirt), and so forth.

The dN300 effect in response to static but not dynamic gestures contrasts with the findings of Ozyurek et al (2007), who report congruency effects elicited by dynamic gestures during both the N300 and N400 time window. One possible explanation for this discrepancy rests in the presentation parameters of the stimuli. In the present experiment, dynamic videos typically included portions of both the preparatory and stroke phases of a gesture, whereas static freeze frames were almost exclusively extracted from the stroke phase. Thus, static items were more similar to the materials used by Ozyurek et al (2007), whose gesture videos included only the stroke phase. Because the meaning of a gesture is thought to be most interpretable during the stroke phase (McNeill, 1992), it is possible that in response to the dynamic gestures in the present study, processes indexed by the N300 were activated slightly later than in response to static items, overlapping with those indexed by the N450. In fact, it is possible that the N300 and N450 do not reflect discrete, serially organized stages of mental activity underlying the comprehension of gestures, but rather concurrent cascading processes.

Support for this idea can be found in Figure 5, which compares isovoltage (left-hand column) and current source density (CSD) maps (right-hand column) of the dN300 and dN450 for both static and dynamic gesture stimuli. Consistent with our statistical analyses of the dN300 and dN450, viz. the absence of a Measurement Interval \times Electrode Site interaction, the isovoltage maps reveal similar scalp topographies for the dN300 and dN450, though the dN450 appears more broadly distributed. Further, the current source density maps indicate that active dipoles within the N300 time window for static gestures were similar to those active during the N450 time window for both gesture types, suggesting similar underlying neural generators were engaged by static and dynamic gestures. Given these commonalities, it is possible that the dN300 static gesture effect reflects the earlier activation of the same processes recruited during the comprehension of dynamic gestures.

Ultimately, conclusive resolution of the question of source localization would require a measurement technique with greater spatial resolution, such as MEG or linear source separation. Nevertheless, similarities in ERP morphology (see Figure 2), topography, and current source density between dynamic and static dN450 suggest that both categories of stimuli likely engaged a highly similar set of neural generators.

4.2 Comparison of Picture and Gesture ERP Effects

Object photos yielded N300 relatedness effects that were marginally larger, and N400 effects that were reliably larger, than static gesture congruity effects measured in the same time windows. This outcome is consistent with the greater effect of relatedness on individuals' subjective ratings of picture pairs than the effect of congruency on ratings of gesture stimuli. As suggested by post-hoc analyses comparing the distribution of ERP effects prompted by these two types of stimuli, the object photo N300 effect was larger over right hemisphere electrode sites than that elicited by static gestures. Visual inspection of isovoltage maps plotting the dN300 and dN400 effects at the point of their maximal magnitude corroborates this claim. Both stimulus types elicit comparably right-lateralized N400 effects; however, the static gesture N300 effect is focused fronto-centrally, whereas the object photo effect is maximal over the anterior right hemisphere (see Figure 6).

Because they are more sensitive to the configuration of current flow (Kutas, Federmeier, & Sereno, 1999), CSD maps (Figure 6) were computed from the isovoltage maps depicted in Figure 6. Comparing within stimulus types, the CSD maps indicate a fairly stable spatial pattern of sources and sinks at the peak of N300 and N400 effects. By contrast, comparison between stimulus types (gestures versus objects) suggests differences in current flow consistent with partially overlapping but non-identical neural generators. Although standard caveats regarding the interpretation of ERP topography effects apply (Urbach & Kutas, 2002), similarities in the ERP relatedness effects elicited by static gestures and photographs of common objects suggest these effects derived from closely overlapping underlying systems.

Observed differences in ERP relatedness effects to objects versus gestures are likely due to the involvement of distinct brain systems for the visual processing of objects and human bodies. Viewing images of the face, body and hands, but not inanimate objects, has been shown to engage cortical regions such as the extrastriate body area (EBA) and occipital and fusiform face areas (FFA) (Bracci, Ietswaart, Peelen, & Cavina-Pratesi, 2010; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Liu, Harris, & Kanwisher, 2009; Op de Beeck, Brants, Baeck, & Wagemans, 2010), as well as the superior temporal sulcus (STS) (Allison, Puce, & McCarthy, 2000; Holle et al., 2008). By contrast, visual features important to object recognition, such as global shape, have been shown to engage distinct cortical regions, such as the lateral occipital complex (LOC) (Grill-Spector, Kourtzi, & Kanwisher, 2001).

Importantly, though, we would like to propose that gestural representations differ from photographic representations not only in their physical instantiation through the human body, but also in the kinds of information they convey. Photographs of an object are successfully recognized as such because they represent visual features that are similar to those experienced when actually seeing that object. By contrast, gestural depictions of an object are meaningful because they represent relations between visuo-spatial features, as in the case of the falling candle. It is precisely this schematic property of depictive gestures that allows them to be used in the metaphorical depiction of concepts that do not possess spatial extent, such as time (Nunez & Sweetser, 2006).

Additionally, depictive gestures can convey meaning through non-iconic, analogically based mappings that are not available in pictures or photographs. In the falling candle gesture, for instance, the speaker's forearm and hand are configured to resemble perceptual features of the candle, such as its length, horizontal orientation, and straight contours. However, an additional mapping is afforded by the analogous structural relationship between the speaker's forearm and hand relative to the candle and the wick. Thus, even though his hand does not particularly resemble a wick, it can nevertheless be construed as such.

In spite of differences in the representational capacities of gestures and pictures, the functional similarities between N300 and N400 effects elicited by these stimulus types, as well as their anterior distributions, also suggest possible commonalities between the systems mediating gesture and picture comprehension. Of course, one concern that might be raised by this line of investigation is whether the present findings generalize to the semantic analysis of co-speech gestures in everyday discourse. Presenting gestures in the absence of the speech that accompanied their production might result in artificial enhancement of attention to gestural forms, for example. Indeed, individual variability in sensitivity to the semantic relations between speech and gesture can impact the degree to which listener comprehension benefits from co-speech iconic gesture (Wu & Coulson, 2010b).

In light of these concerns, it is important to keep in mind that the present experimental paradigm affords investigation of the neurocognitive processes invoked to interpret iconic

gestures, independently of the processes involved in speech-gesture integration. This line of investigation is in keeping with a number of studies demonstrating that iconic gestures are interpretable even when concurrent speech is absent or obscured (Obermeier, Dolk, & Gunter, 2011; Rogers, 1978; Wu & Coulson, 2007b). Further, in natural settings, listeners have occasion to interpret iconic gestures when the supportive context of concurrent speech is significantly diminished – as in the case of gestures produced during pauses as a speaker searches for a particular word, or during conversation in a noisy environment that makes speech difficult to discern. Finally, the presentation of gestures without concurrent speech was warranted in order to make them more comparable to the object photographs examined in Experiment 2.

A second consideration engendered by the experimental paradigm used here is whether findings obtained in response to static representations generalize to the interpretation of gestures in face-to-face settings. One line of evidence suggesting static and dynamic representations impact comprehension in similar ways is our finding that both static and dynamic gestures served as effective primes for related probe words. The N400 relatedness effect (reduced amplitude N400 for related, e.g. “flowerpot”, relative to unrelated, e.g. “drainpipe”, probe words) was similar in size and topography after static and dynamic gestures (see section 3.5).

Similarities between the processing of static and dynamic gestures is also in keeping with a number of studies demonstrating that static representations elicit brain response similar to that prompted by their dynamic counterparts. For instance, photographs depicting implied motion elicit activation in brain regions implicated in the visuo-perceptual analysis of motion, such as MT/MST (Kourtzi & Kanwisher, 2000). Analogously, both photographic and video depictions of hand grasps result in similar activation of action comprehension systems (as reflected by reduced power in the mu frequency band (around 10.5 and 21 Hz) over somatomotor cortex) (Perry & Bentin, 2009; see also Pineda & Hecht, 2009).

5. Conclusion

The present study has confirmed and advanced existing cognitive neuroscience research on the real-time comprehension of gestures. In keeping with previous findings (Wu & Coulson, 2005), ERPs time locked to the onset of depictive gestures elicited a broadly distributed negative component – the N450 – with reduced amplitude for congruent as compared to incongruent items. The gesture N450 was hypothesized to index the semantic integration of gestures with information made available in preceding cartoons, in keeping with the N400 observed in picture priming paradigms. Additionally, contextually incongruent static gestures resulted in enhanced N300 relative to congruent items. We interpreted this finding as evidence that the process of categorizing percepts activated by static gesture stills is affected by the representational properties of the gestures. In this way, depictive hand configurations are similar to the contours and shapes that allow pictures of objects to be successfully recognized. Observed differences in the topographic distribution of the gesture dN450 and the object dN400 are likely related both to the involvement of brain systems specialized for visual analysis of the human body, and to the importance of more abstract mappings and relational structure in the attribution of meaning to depictive gestures.

Highlights

We compared neuro-cognitive processes engaged by pictures and depictive gestures.

Prime-target trial types: photo-photo, cartoon-gesture video, cartoon- gesture jpeg.

Targets were congruent or incongruent with context primes.

ERPs time-locked to all three target types exhibit N400 congruency effect.
ERPs to photos and gesture freeze frames exhibit N300 congruency effect as well.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by grants to SC from the NSF (#BCS-0843946), an Innovative Research Award from the Kavli Institute for Brain & Mind, and an NIH Training Grant (T32 MH20002-10) award to YW.

Works Cited

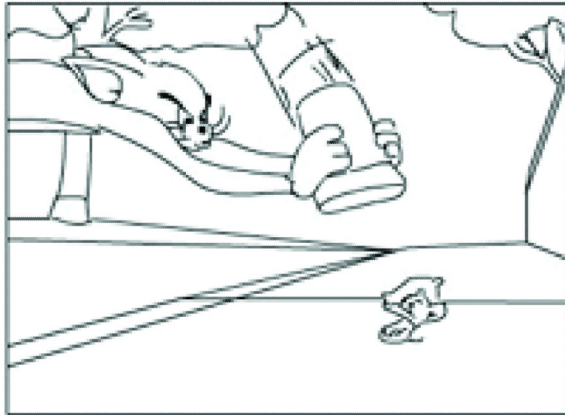
- Allison T, Puce A, McCarthy G. Social perception from visual cues: Role of the STS. *Trends in Cognitive Sciences*. 2000; 4(7):267–278. [PubMed: 10859571]
- Barrett SE, Rugg MD. Event-related potentials and the semantic matching of pictures. *Brain and Cognition*. 1990; 14:201–212. [PubMed: 2285513]
- Bracci S, Ietswaart M, Peelen MV, Cavina-Pratesi C. Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*. 2010; 103(6):3389–3397. [PubMed: 20393066]
- Buccino G, Binkofski F, Riggio L. The mirror neuron system and action recognition. *Brain and Language*. 2004; 89:370–376. [PubMed: 15068920]
- Corballis MC. From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*. 2003; 26:199–260. [PubMed: 14621511]
- Critescu TC, Nobre AC. Differential modulation of word recognition by semantic and spatial orienting of attention. *Journal of Cognitive Neuroscience*. 2008; 20(5):787–801. [PubMed: 18201125]
- Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N. Domain specificity in visual cortex. *Cerebral Cortex*. 2006; 16(10):1453–1461. [PubMed: 16339084]
- Federmeier KD, Kutas M. Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:202–224.
- Federmeier KD, Kutas M. Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*. 2002; 40:730–747. [PubMed: 11900725]
- Feyereisen, P.; de Lannoy, JD. *Gestures and speech: Psychological investigations*. New York, NY: Cambridge University Press; 1991.
- Ganis G, Kutas M, Sereno MI. The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*. 1996; 8(2):89–106.
- Geisser S, Greenhouse S. On methods in the analysis of profile data. *Psychometrika*. 1959; 24:95–112.
- Graham JA, Argyle M. A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*. 1975; 10:57–67.
- Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. *Vision Research*. 2001; 41:1409–1422. [PubMed: 11322983]
- Gunter TC, Bach P. Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters*. 2004; 372:52–56. [PubMed: 15531087]
- Hamm JP, Johnson BW, Kirk IJ. Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*. 2002; 113:1339–1450. [PubMed: 12140015]
- Holcomb PJ. Automatic and attentional processing: An event-related brain potential analysis of semantic processing. *Brain and Language*. 1988; 35:66–85. [PubMed: 3179703]

- Holcomb PJ, Kounios J, Anderson JE, West CW. Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1999; 25(3):721–742.
- Holcomb PJ, McPherson WB. Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*. 1994; 24:259–276. [PubMed: 8185897]
- Holle H, Gunter TC, Ruschemeyer S-A, Hennenlotter A, Jacoboni M. Neural correlates of the processing of co-speech gestures. *NeuroImage*. 2008; 15(39):2010–2024. [PubMed: 18093845]
- Jacoboni M. Imitation, empathy, and mirror neurons. *Annual Review of Psychology*. 2009; 60:653–670.
- Kelly SD, Creigh P, Bartolotti J. Integrating speech and iconic gestures in a stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*. 2009; 22(4):683–694. [PubMed: 19413483]
- Kelly SD, Manning SM, Rodak S. Gesture gives a hand to language and learning: perspectives from Cognitive Neuroscience, Developmental Psychology, and Education. *Language and Linguistics Compass*. 2008; 2(4):569–588.
- Kelly SD, Ozyurek A. Gesture, brain, and language. *Brain and Language*. 2007; 101(3):181–184. [PubMed: 17433429]
- Kelly SD, Ward S, Creigh P, Bartolotti J. An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*. 2007; 101:222–233. [PubMed: 16997367]
- Kendon, A. Some relationships between body motion and speech. In: Siegman, A.; Pope, B., editors. *Studies in dyadic communication*. New York: Pergamon Press; 1972. p. 177-210.
- Kendon, A. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press; 2004.
- Kounios J, Holcomb PJ. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1994; 20(4):804–823.
- Kourtzi Z, Kanwisher N. Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*. 2000; 12(1):48–55. [PubMed: 10769305]
- Kutas, M.; Federmeier, KD.; Sereno, MI. Current approaches to mapping language in electromagnetic space. In: Brown, CM.; Hagoort, P., editors. *The Neurocognition of language*. Oxford University Press; 1999. p. 359-392.
- Kutas M, Hillyard SA. Event-related potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*. 1980; 11:99–116. [PubMed: 7272388]
- Kutas M, Hillyard SA. Brain potentials during reading reflect word expectancy and semantic association. *Nature*. 1984; 307:161–163. [PubMed: 6690995]
- Lim VK, Wilson AJ, Hamm JP, Phillips N, Iwabuchi SJ, Corballis MC, et al. Semantic processing of mathematical gestures. *Brain and Cognition*. 2009; 71:306–312. [PubMed: 19665831]
- Liu J, Harris A, Kanwisher N. Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*. 2009; 22(1):203–211. [PubMed: 19302006]
- McNeil N, Alibali MW, Evans JL. The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*. 2000; 24(2):131–150.
- McNeill, D. *Hand and mind*. Chicago: Chicago University Press; 1992.
- McPherson WB, Holcomb PJ. An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*. 1999; 36:53–65. [PubMed: 10098380]
- Mudrik L, Lamy D, Deouell LY. ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*. 2010; 48:507–517. [PubMed: 19837103]
- Nunez RE, Sweetser E. With the future behind them: Convergent evidence from Aymara language and gesture in crosslinguistic comparison of spatial construals of time. *Cognitive Science*. 2006; 30(3): 401–450. [PubMed: 21702821]
- Nuwer MR, Comi G, Emerson R, Fuglsang-Frederiksen A, Guerit JM, Hinrichs H, et al. IFCN standards for digital recording of clinical EEG. *The international federation of clinical neurophysiology. Electroencephalography and Clinical Neurophysiology Supplement*. 1999; 52:11–14. [PubMed: 10590972]

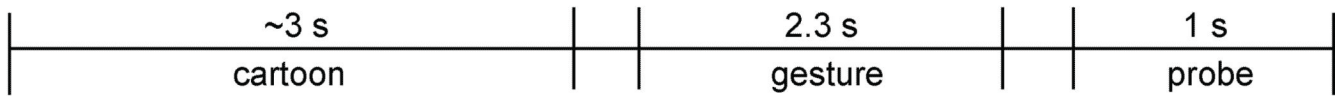
- Obermeier, C.; Dolk, T.; Gunter, TC. *Cortex*. 2011. The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. in press.
- Oldfield RC. The assesment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 1971; 9:97–113. [PubMed: 5146491]
- Op de Beeck HP, Brants M, Baeck A, Wagemans J. Distributed subordinate specificity for bodies, faces, and buildings in human ventral visual cortex. *NeuroImage*. 2010; 49(4):3414–3425. [PubMed: 19922804]
- Ozyurek A, Willems RM, Kita S, Hagoort P. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*. 2007; 19(4):605–616. [PubMed: 17381252]
- Paulmann S, Jessen S, Kotz SA. Investigating the multimodal nature of human communication: Insights from ERPs. *Journal of Psychophysiology*. 2009; 23(2):63–76.
- Perry A, Bentin S. Mirror activity in the human brain while observing hand movements: A comparison between EEG desynchronization in the mu-range and previous fMRI results. *Brain Research*. 2009; 1282:126–132. [PubMed: 19500557]
- Pineda JA, Hecht E. Mirroring and mu rhythm involvement in social cognition: Are there dissociable subcomponents of theory of mind? *Biological Psychology*. 2009; 80(3):306–314. [PubMed: 19063933]
- Rizzolatti G, Arbib MA. Language within our grasp. *Trends in Neuroscience*. 1998; 21:188–194.
- Rizzolatti G, Craighero L. The mirror-neuron system. *Neuroscience*. 2004; 27:169–192.
- Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*. 1996; 3:131–141. [PubMed: 8713554]
- Rogers WT. The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research*. 1978; 5:54–62.
- Schendan HE, Kutas M. Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*. 2002; 40:931–945. [PubMed: 11900745]
- Schendan HE, Kutas M. Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience*. 2003; 15(1):111–135. [PubMed: 12590847]
- Schendan HE, Kutas M. Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*. 2007; 19(5):734–749. [PubMed: 17488201]
- Schendan HE, Maher SM. Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *NeuroImage*. 2009; 44:1423–1438. [PubMed: 19010426]
- Sitnikova T, Kuperberg G, Holcomb PJ. Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*. 2003; 40(1):160–164. [PubMed: 12751813]
- Urbach TP, Kutas M. The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology*. 2002; 39:791–808. [PubMed: 12462507]
- van Berkum JJ, Hagoort P, Brown CM. Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience*. 1999; 11(6):657–671. [PubMed: 10601747]
- West WC, Holcomb PJ. Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*. 2002; 13(3):363–375. [PubMed: 11919001]
- Willems RM, Ozyurek A, Hagoort P. When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*. 2007; 17:2322–2333. [PubMed: 17159232]
- Willems RM, Ozyurek A, Hagoort P. Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*. 2008; 20(7): 1235–1249. [PubMed: 18284352]
- Willems RM, Ozyurek A, Hagoort P. Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*. 2009; 47:1992–2004. [PubMed: 19497376]
- Wu YC, Coulson S. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*. 2005; 42(6):654–667. [PubMed: 16364061]

- Wu YC, Coulson S. How iconic gestures enhance communication: An ERP study. *Brain and Language*. 2007a; 101(3):234–245. [PubMed: 17222897]
- Wu YC, Coulson S. Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*. 2007b; 14(1):57–63. [PubMed: 17546731]
- Wu YC, Coulson S. Gestures modulate speech processing early in utterances. *NeuroReport*. 2010a; 21(7):522–526. [PubMed: 20375745]
- Wu, YC.; Coulson, S. Iconic gestures facilitate word and message processing: The Multi-Level Integration model of audio-visual discourse comprehension. Poster presented at the Neurobiology of Language Conference; Rancho Bernardo California. 2010b.

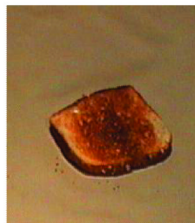
A



flowerpot



B



kitchen

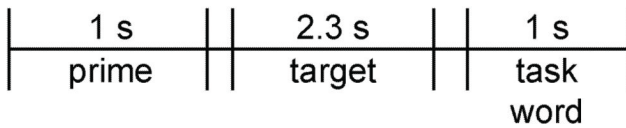


Figure 1. Experiment Procedures

A. Experiment 1: Trials consisted of a short cartoon segment followed by either a dynamic or static gesture and then a probe word. **B.** Experiment 2: Trials consisted of a context photograph, a photograph that was either related or unrelated to the context, or was unidentifiable, and then a probe word.

Dynamic Gestures

Static Gestures

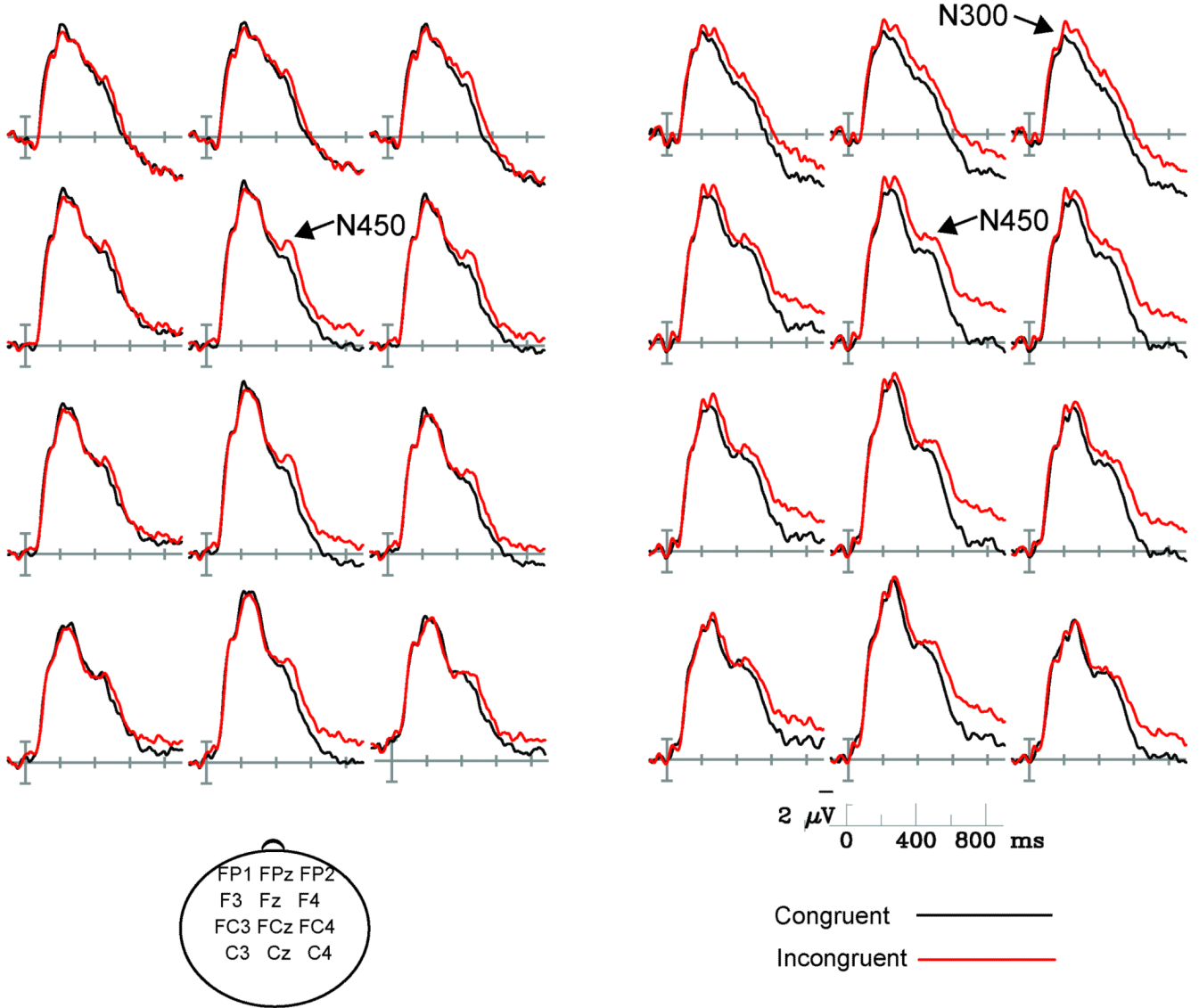


Figure 2. ERPs time-locked to the onset of gesture videos and static gesture stills
 Both types of stimuli elicited a reduced N450 when primed by a congruent as opposed to incongruent context. Contextually congruent static gesture frames also elicited an N300 component of reduced amplitude relative to incongruent ones. Note that negative voltage is plotted up in this and all subsequent figures.

Common Objects

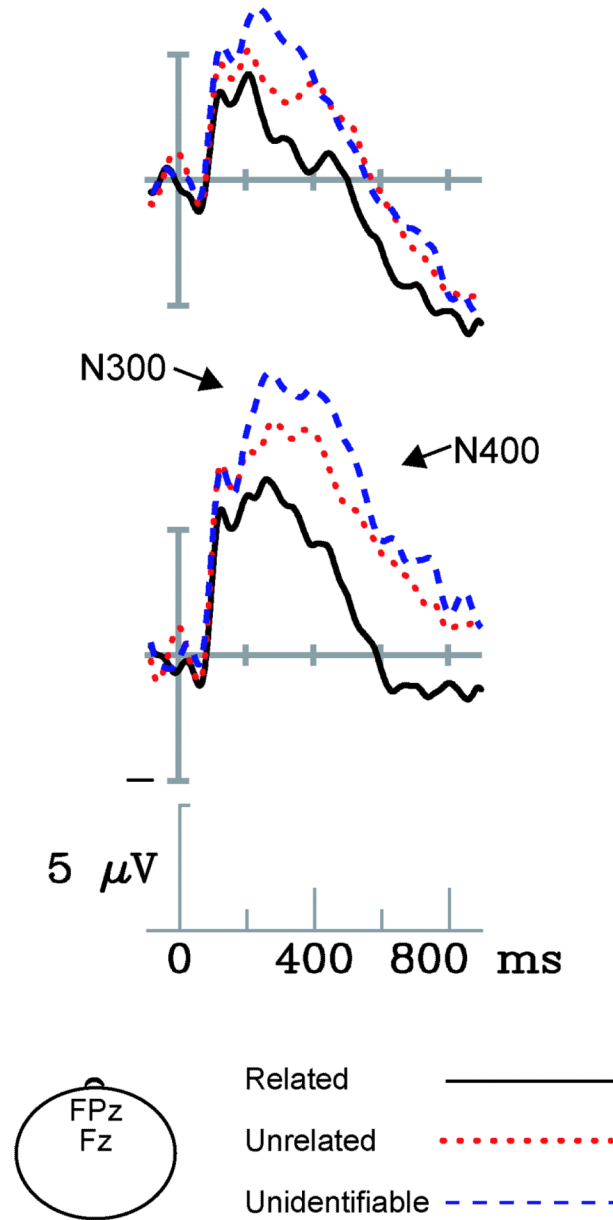


Figure 3. ERPs elicited by related, unrelated, and unidentifiable photographs of common objects
Targets elicited attenuated N300 and N400 components when occurring in related versus unrelated picture pairs. Unrelated targets elicited less negative ERPs within the N300 and N400 time window relative to unidentifiable objects.



Figure 4. Example from a cartoon prime (artist's depiction) and the subsequent static gesture used in the experiment

Cartoons and gestures bear primarily relational rather than perceptual similarities.

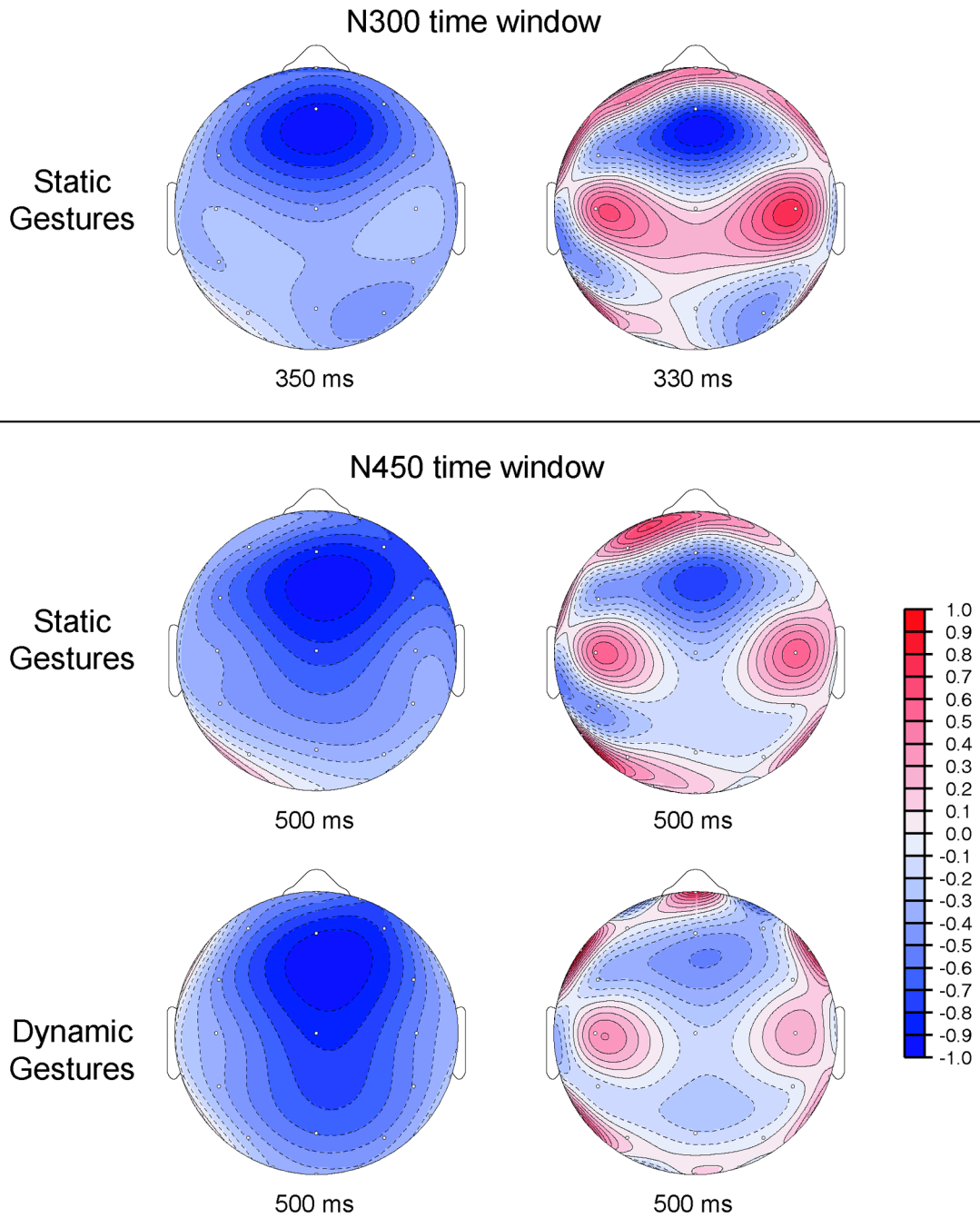


Figure 5. The scalp distribution of the gesture N300 and N450 effect
 Normalized interpolated isovoltage maps (left) and current source density maps (CSD) (right) were computed from congruity effects elicited by dynamic and static gestures at the peak value of the congruity effect within the N300 (top row) and N450 time windows (bottom 2 rows) post-stimulus onset. (Activity elicited by dynamic gestures is not shown for the N300 time window because congruity effects in those conditions did not start until 400 ms post-stimulus.)

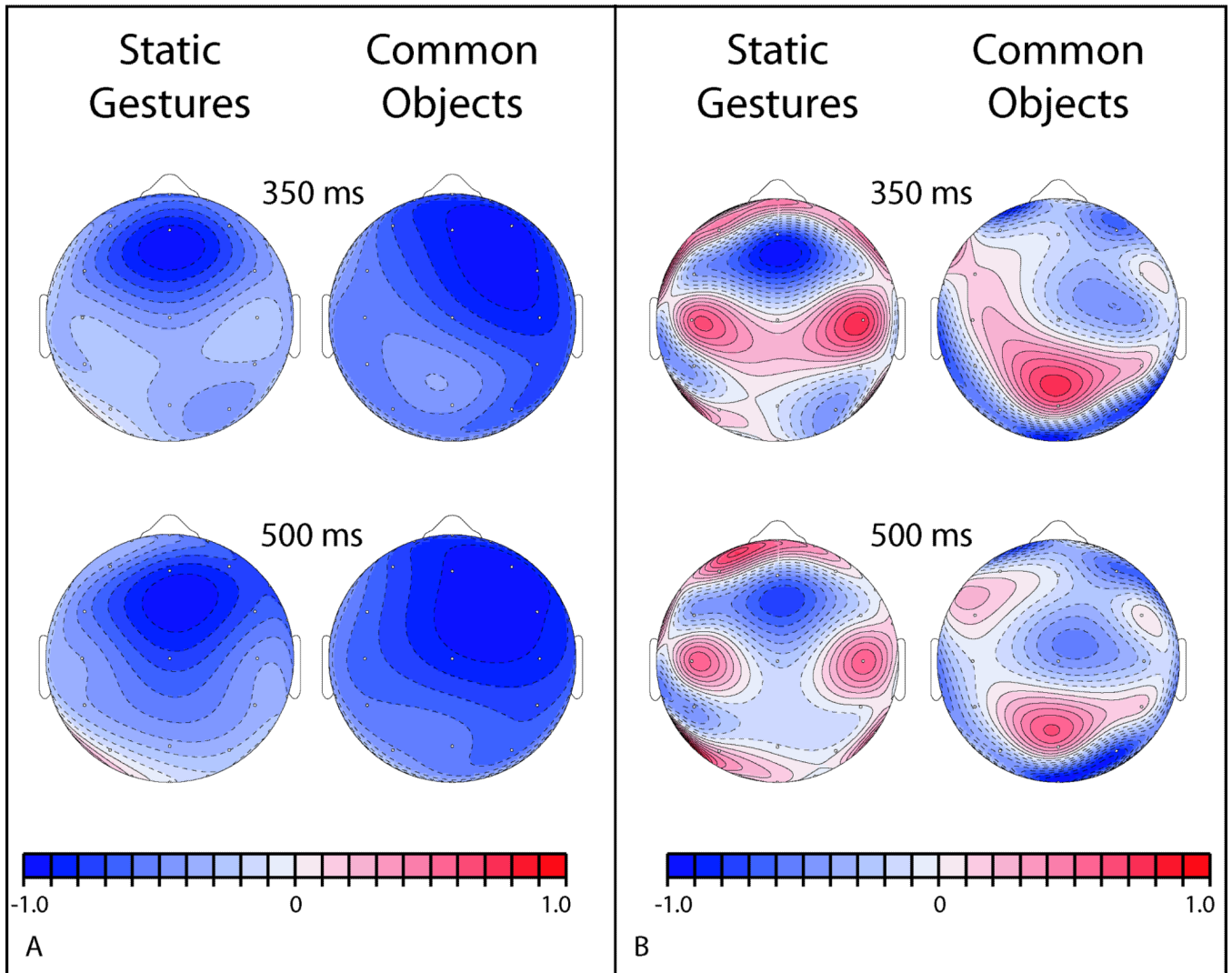


Figure 6. The scalp distribution of semantic mismatch effects elicited by static gestures and photographs of common objects

A. The topography of the gesture congruity effect (left column) was computed at 350 ms (dN300) and 500 ms (dN450) post-stimulus. The topography of the photo relatedness effect (right column) was computed at 350 ms (dN300) and 400 ms (dN400) post-stimulus. At both time points, the topography of the congruity effect for static gestures differs from that for common objects. **B.** CSD maps (right panel) were derived by computing the second spatial derivative of each difference wave (left panel). At both time points, visual inspection suggests a different pattern of sources and sinks for static gestures and common objects.