Psychophysiology, 42 (2005), 654–667. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2005.00356.x

Meaningful gestures: Electrophysiological indices of iconic gesture comprehension

YING CHOON WU AND SEANA COULSON
Cognitive Science Department, University of California, San Diego, La Jolla, California, USA

Abstract

To assess semantic processing of iconic gestures, EEG (29 scalp sites) was recorded as adults watched cartoon segments paired with soundless videos of congruous and incongruous gestures followed by probe words. Event-related potentials time-locked to the onset of gestures and probe words were measured in two experiments. In Experiment 1, participants judged the congruency between gestures and cartoons. Gestures elicited an N400-like component (gesture N450) that was larger for incongruent than congruent items, as well as a late positivity that was larger for congruent items. In Experiment 2, participants assessed the relatedness between probe words and preceding cartoon-gesture pairs. N450 effects to gestures were observed without overlapping positivity. These findings suggest that iconic gestures are subject to semantic processes analogous to those evoked by other meaningful representations, such as pictures and words.

Descriptors: Gesture, N400, Conceptual integration, Semantics

Although it has long been noted that people produce rhythmic movements of their hands and arms as they speak, the communicative significance of these movements is not well understood (Krauss, 1998). Until recently, this issue has been studied mainly by researchers in ethnography and cognitive psychology. Such research suggests that co-speech gestures may improve communicative coordination in a variety of ways. Gestures have been shown, for example, to direct attention (Goodwin, 2000), modulate speech acts (Kendon, 2000), and, of particular importance to the present investigation, to illustrate elements of the speaker's conceptual world. For example, McNeill (1992) shows how in one class of gesture, called *iconic* or *physiographic* (Efron, 1972), speakers typically move their hands and arms to create a dynamic visual representation of semantic properties related to the content of their speech. A speaker might demonstrate the shape of a platter, for instance, by tracing an oval in the air. Here we consider whether these sorts of iconic gestures are subjected to semantic processing by listeners.

McNeill (1992) has theorized that gesture and speech constitute opposed, but complementary, dimensions of thought, with gestures expressing holistic, imagistic relations, and speech expressing linear, componentially segmentable ones. In this view, iconic co-speech gestures are likely to provide additional semantic information about the content of the talk in progress, helping listeners to build an enriched conceptual representation of the speaker's message. Behavioral findings in support of this hypothesis have been obtained in two distinct types of experimental paradigms. In one approach, measures of comprehension are compared from individuals exposed to speech in either an audio-only medium or in video form, with accompanying gestures visible. Several studies using this technique have found that when the speakers' accompanying gestures are visible, listeners are better able to comprehend the sizes, locations, category membership, agency, and action type of described events and objects (Beattie & Shovelton, 1999a, 1999b, 2002; Rogers, 1978).

In another approach, the impact of co-speech gestures is indexed by confusions produced by materials in which gestures and speech convey conflicting information. Listeners are typically asked to retell or evaluate a speaker's description of an event, and evidence of sensitivity to gesturally transmitted information is gauged in the content of the listeners' own verbal and gestural responses (Alibali, Flevares, & Goldin-Meadow, 1997; Cassell, McNeill, & McCullough, 1999; Goldin-Meadow & Sandhofer, 1999; Kelly, Barr, Church, & Lynch, 1999). In one such experiment, gestures that did not correspond with speech precisely, as in the case of an actor making a punching gesture while saying, "whacks him one," resulted in accounts of the narrative that reflected contributions from both sources (e.g., "And Granny like punches him or something and you know he whacks him . . ."; Cassell et al., 1999, p. 20). These findings suggest that at least in some contexts, iconic gestures engage semantic processes and can produce measurable effects on observer comprehension.

On the other hand, some researchers have argued that co-speech gestures are minimally communicative and are generated as epiphenomena of speech production processes. In this view, iconic co-speech gestures serve to benefit the speaker by facilitating lexical access, but have minimal impact on the listener. In a study that

---

The authors thank Marta Kutas and three anonymous reviewers for suggestions on an earlier version of the manuscript. Y.W. was supported by Institute for Neural Computation NIH Training Grant #MH 2002-07 and Center for Research on Language NSF Training Grant DC 000-16.

Address reprint requests to: Ying Choon Wu, Cognitive Science, Dept. 0515, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA. E-mail: ywu@cogsci.ucsd.edu.

compared the comprehension of speech alone as opposed to speech with concurrent gestures, no effect of gesture visibility was found on overhearers' abilities to identify abstract designs, synthesized sounds, or flavors of tea after exposure to videotaped vignettes in which a speaker spontaneously described the target objects (Krauss, Dushay, Chen, & Rauscher, 1995). Other behavioral studies suggest that listeners rely heavily on semantic information conveyed through speech in order to attribute meaning to accompanying gestures (Krauss, Morell-Samuels, & Colasante, 1991).

However, one limitation of this prior research is the off-line nature of the dependent measures. Assessing the effects of gesture on subsequent comprehension affords only an indirect view of the cognitive activity evoked by the gestures themselves. Because such approaches have yielded mixed results, real-time measurement techniques, such as ERPs, are critical for understanding the effects of co-speech gestures on comprehension.

In one relevant study, participants' ERPs were time-locked to the utterance of single words accompanied by either congruent or incongruent gestures (Kelly, Kravitz, & Hopkins, 2004). Stimuli were constructed by videotaping an actor as he gestured to either a tall, thin glass or a short, wide dish in front of him while saying one of four speech tokens—namely, *tall*, *thin*, *short*, or *wide*. Gestures indicated the location of these two items and also depicted either the height or width of their referent. Speech tokens were presented either without accompanying gestures or were presented with matching, entirely mismatching, or complementary gestures. Relative to the other conditions, mismatching trials elicited consistently more negative ERPs between 324 and 648 ms at bilateral temporal electrode sites. These findings suggest that incongruent concurrent gestures can negatively affect the processing of speech; however, it is still unknown whether congruent co-speech gestures facilitate comprehension.

Moreover, it is largely unknown how gestures themselves are processed (though see Gunter & Bach, 2004, for an ERP study investigating the comprehension of conventionalized hand signs known as emblems). If it is correct that information encoded in speech and gesture is integrated in comprehension, we would expect manipulations of gesture congruency to affect not only brain response to speech, as demonstrated by Kelly et al. (2004), but to gestures as well. Support for this prediction would constitute necessary, though not sufficient, evidence for the speech–gesture integration hypothesis. Alternatively, if co-speech gestures affect the processing of speech, but are subject to only minimal semantic analysis, as suggested by Krauss et al. (1995, 1991), no effects of congruency on brain responses to gestures are expected.

The present study addresses the semantic impact of gesture by recording ERPs as participants watch video clips of a speaker's spontaneously produced iconic gestures. Our stimuli came from a corpus of iconic, co-speech gestures that was collected by videotaping an individual describing cartoon segments. He was told that the experimenters were creating stimuli for a subsequent memory experiment and was unaware of the intent to elicit spontaneous gestures. To create a set of congruous and incongruous cartoon–gesture pairs, occurrences of co-speech iconic gesture were digitized into short video clips and paired either with the original cartoon clips utilized in their elicitation or with clips that elicited different gestures. Because the accompanying speech in these clips contained significant cues to their congruity with the preceding cartoons, the gestures in this study were presented as soundless video clips. This enabled us to test whether gestures can affect comprehension in the absence of other sources of semantic input, and to assess whether iconic gestures undergo semantic processing.

We hypothesize that the integration of iconic gestures with contextually activated knowledge is mediated by some of the same semantic integration processes engaged during the comprehension of more uncontroversially meaningful image-based stimuli, such as pictures or line drawings. This hypothesis can be tested given previous findings of electrophysiological correlates of the semantic analysis of images. For example, incongruous prime–target picture pairs have been shown to elicit an anterior negativity peaking around 300 ms after the onset of the stimulus (N300), as well as a more broadly distributed negativity peaking approximately 400 ms post-stimulus (N400; Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999).

The discovery of an N400 response to incongruous images has led to the suggestion that the neural system involved in picture comprehension may function similarly to the system responsible for the "classic" N400 elicited by linguistic stimuli. A well-studied ERP component, the lexical N400 is thought to reflect certain aspects of meaning processing (Kutas & Federmeier, 2000). The last word of a sentence that ends as expected typically elicits little or no N400, whereas unexpected sentence completions elicit an N400 component with a large amplitude (Kutas & Hillyard, 1980, 1984). N400 amplitude is also sensitive to intermediate levels of semantic constraint such that it can be interpreted as an index of the degree to which a word fits its context (Kutas & Hillyard, 1984).

Like its lexical counterpart, the picture N400 also exhibits sensitivity to different degrees of relatedness, being larger for items that are moderately related than for highly related ones (Kutas & Hillyard, 1988; McPherson & Holcomb, 1999). Moreover, just as pseudo-words elicit larger N400s than unrelated words (Holcomb, 1988), unrecognizable images elicit larger N400s than recognizable ones (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999). Further, the amplitude of both the word and the picture N400 is modulated by the global, discourse-level coherence of a word (van Berkum, Hagoort, & Brown, 1999) or picture (West & Holcomb, 2002) within a story context. Because N400 effects elicited by pictures tend to be larger over the front of the head, whereas lexical N400 effects tend to be largest centro-parietally, it is unlikely that wholly identical systems mediate word and picture comprehension (Ganis, Kutas, & Sereno, 1996). However, given the similar time course and sensitivity to preceding context shared by the lexical and picture N400, the comprehension of both words and pictures appears to involve neural systems that function in a comparable manner.

By contrast, the N300, a negative peak in the ERP elicited by pictures and photographs of common objects, has been argued to index processes specific to image comprehension. Like the N400, the N300 is modulated by contextual congruity (Hamm, Johnson, & Kirk, 2002). Yet, in a study involving image pairs with graded degrees of associative relatedness (highly related, moderately related, and unrelated), the amplitude of the N300 reflected differentiation only for related and unrelated items, but not for moderately and highly related ones. Further, N300 effects tend to be largest over anterior electrode sites, whereas N400 effects are more broadly distributed (McPherson & Holcomb, 1999).

These differences support the view that the picture N300 and N400 reflect different aspects of image comprehension. A variety of studies have demonstrated that the N300 is modulated by the difficulty of mapping perceptual input onto stored semantic representations. Fragmented line drawings of objects that cannot be named, for example, elicit enhanced N300 relative to identifiable fragmented items (Schendan & Kutas, 2002). Further, contextually incongruent pictures that share basic-level features with the expected target (i.e., within-category violations such as a donkey instead of a zebra, or a collie instead of a poodle) result in reduced N300 relative to between-category violations (a dalmation instead of a zebra or a collie instead of a mallard; Federmeier & Kutas, 2002; Hamm et al., 2002). On the basis of these studies, the N300 has been proposed to index the process whereby image-based representations in long-term memory are accessed as a result of the structural analysis of perceptual input (Schendan & Kutas, 2002; West & Holcomb, 2002). In contrast, the N400 family of potentials is thought in general to index brain activity mediating the integration of semantic activations triggered by a current event with those prompted by previous ones (Kutas & Hillyard, 1984).

If the integration of semiotic features of gestures with contextually active information recruits integration processes similar to those engaged during picture and language comprehension, we might expect manipulations of gesture congruency to result in N400 effects. Moreover, if the time course of gesture comprehension is similar to that of picture comprehension, discernible N300 effects may also be observed. In Experiment 1, we tested these predictions by measuring ERPs elicited by gestures as participants indicated whether or not they were congruent with their preceding cartoon context. In Experiment 2, we utilized the same set of stimuli, but employed a task that did not require overt semantic analysis of gestures.

If N400-like congruency effects are elicited by gestures, these experiments would provide real-time processing evidence that iconic gesture comprehension recruits semantic integration processes analogous to those involved in understanding other kinds of contentful representations, such as words and pictures. N300 effects obtained in response to gestures would demonstrate that this component does not index processes specific to the analysis of static images, but rather, mediates the comprehension of dynamic ones as well.

EXPERIMENT 1

Method

Participants

Seventeen volunteers were paid $24 or awarded course credit for participation. The data of 5 participants were excluded because of excessive artifacts, including mainly eye movements (greater than 40% of trials in critical bins). The remaining 12 individuals (6 women and 6 men; mean age = 21.5 years) were healthy, right-handed, fluent English speakers with no history of neurological impairment. Their mean laterality quotient, as assessed by the Edinburgh Inventory (Oldfield, 1971), was .725, indicating a fairly strong bias toward right-handedness.

Materials

Stimuli were 160 pairs of cartoon and gesture video clips. Cartoon clips were derived by digitizing popular television cartoon shows (e.g., *Tom and Jerry*, *Daffy Duck*, *The Roadrunner*) into short, soundless segments with Speed Razor software. On average, cartoons lasted 3 s, and typically depicted one or two salient actions or events (e.g., Nibbles jabs Tom's foot and offers him a firecracker; a rock rolls toward the Roadrunner; Jerry rings a bell).

To construct the gesture clips, a naive individual was videotaped using a Sony Hi 8 video recorder while describing these cartoon segments. Three recording sessions took place. He was told that his videotaped speech would be utilized in the construction of stimuli for a memory experiment, and was instructed to describe each clip in as much detail as possible; however, no mention of gestures was made. Spontaneous gestures that were judged to represent elements within the corresponding cartoons were digitized into soundless segments of 48 frames each. Typically, the first frame coincided with the onset of the stroke phase of each gesture. In fewer than 9% of trials, the image sequence began in the preparation phase (e.g., the prestroke hold), primarily in cases where the stroke was executed very quickly. The presentation of each set of gesture frames lasted 2.3 s. On average, within each set of frames, gesture production extended for 2 s ($SD = 336$ ms). In 62% of trials, gesture production continued until the final frame. Gestures typically either reenacted actions performed in the cartoon from a first person perspective (turning a doorknob, swinging a bat, lowering a rope) or depicted salient features of an event (the path of a careening rock, the speed of falling apples) or object (the shape of a platter, the orientation of truck bed) (see Table 1). In some cases, gestures highlighted central relations depicted in the cartoon; in others, they emphasized fairly incidental details.

Congruous trials were those in which cartoon clips were paired with the original gestures produced while the narrator described them. Incongruous trials involved mismatches. A normative study was conducted to ensure the generalizability of experimenter intuitions about congruency relations between cartoons and gestures. Ten individuals subjectively rated the degree of relatedness between cartoons and gestures on a scale of 1 to 5, with 5 designating the highest level. The average relatedness rating was 4.2 ($SD = 0.16$) for congruous trials and 1.3 ($SD = 0.22$) for incongruous ones.

Two lists were constructed, each containing 80 congruous and 80 incongruous trials. No cartoon or gesture clip was repeated on either list, but across lists, each gesture appeared once as a congruous stimulus and once as an incongruous one.

Procedure

Trials began with a fixation cross, presented in the center of a 17-in. color monitor. The cartoon and gesture clips were presented

Table 1. *Types of Iconic Gestures Used as Stimuli*

	Depictive properties		
	Action schemas	Features of events	Features of objects
Quantity	85	41	34
Example	unlocking a door applying glue adjusting a robot lifting a lid	path of an arrow in flight chomping jaws impact of a collision a dog striding	length of a bridge shape of a panel location of buttons shape of a lid

Total = 160.

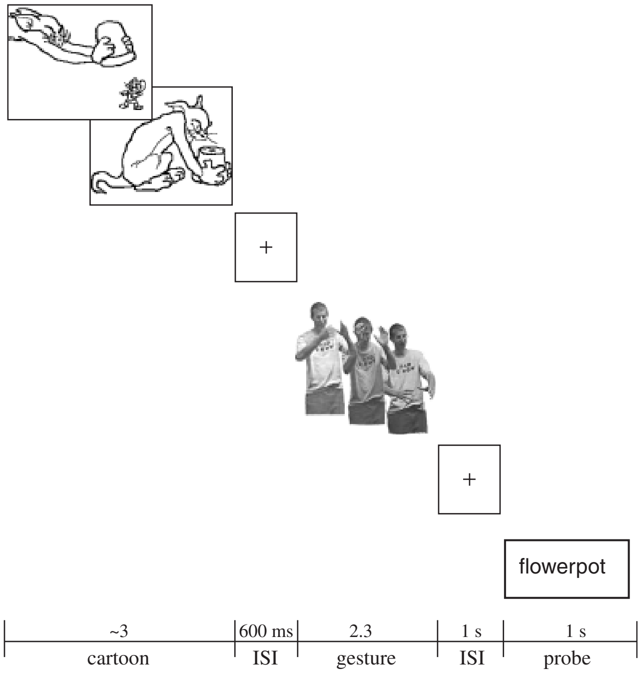

Figure 1. Sample trial: A short cartoon segment was followed by a congruent or incongruent gesture video and then a probe word.

at a rate of 48 ms per frame with a 600-ms pause before the onset of the gesture (to allow participants time to establish central fixation). Although cartoons varied in length (mean = 2949 ms, *SD* = 900 ms), 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 for 1 s (see Figure 1). Participants were not required to make any behavioral response to probe words, which were being piloted for Experiment 2. A short pause (~ 5–6 s) 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 (the speaker himself subtended approximately 3°–6° horizontally and 6.8° vertically). Primarily the head, arms, and upper torso of the speaker were visible in each gesture trial.

Participants were told that they would watch a series of cartoon segments, each followed by video clips of a man describing either the immediately preceding cartoon or a different one. They were asked to press YES or NO on a button box as soon as they felt confident that his description matched or did not match the preceding cartoon. Response hand was counterbalanced across subjects. Figure 1 shows a schematic of a sample trial. Four additional trials were used in a practice block at the outset.

EEG Recording

The electroencephalogram (EEG) was recorded using tin electrodes at 29 standard International 10–20 sites (Nuwer et al., 1999), including 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 channels (F7, FT7, TP7, T5, F8, FT8, TP8, T6). Electrodes were also placed on the right mastoid for off-line rereferencing, below the right eye for monitoring blinks, and at the outer canthi for monitoring eye movements. All electrodes were referenced on-line to the left mastoid, and impedences maintained below 5 kΩ. 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.

Behavioral Data Analysis

Participants' mean accuracy and response latencies were assessed with repeated-measures ANOVA with both subjects (F_1) and items (F_2) as random variables. Analyses were conducted on responses occurring within a 3-s window after stimulus onset (5% congruous and 8% incongruous trials lost due to trimming), and a 4-s window (only 0.02% congruous and 0.04% incongruous trials lost).

EEG Analysis

Trials affected by artifacts such as blinks, eye movements, blocking, and drift were rejected off-line by automated routines whose thresholds were optimized for each data set. Blinks were indexed by the difference in voltage measured at the lower eye electrode and FP2. Cases in which this comparison exceeded approximately $\pm 16\ \mu V$ were flagged for contamination from blinking. Eye movements were monitored by means of a bipolar montage of electrodes affixed to the outer canthi. On average, epochs in which the difference between the maximum and the minimum values was greater than approximately $\pm 6\ \mu V$ were flagged for contamination from eye movements.

Artifact-free ERP averages time-locked to the onset of gestures were constructed from 300 ms before stimulus onset to 2760 ms after. Because effects comparable to those discussed in extant research on picture comprehension occurred before 1200 ms poststimulus, only analyses within 1200 ms are reported. Trials accurately categorized by participants were sorted and averaged. Those that elicited inaccurate responses were excluded. On average, the congruent bin contained 43 trials (40 median), and the incongruent bin 46 trials (44 median). The mean artifact rejection rate was 32% ($SD = 20\%$) for congruous items and 35% ($SD = 20\%$) for incongruous ones. A two-tailed matched pairs t test indicated that the difference in artifact rejection rates between these conditions was not reliable, $t(11) = 1.47$, $p = .168$). This relatively high artifact rejection rate resulted mainly from eye movements during the presentation of gestures.

Congruency effects were assessed by measuring the mean amplitude (relative to the prestimulus baseline) and peak latencies of ERPs time-locked to gesture onset from 300 to 400 ms, 400 to 600 ms, 600 to 900 ms, and 900 to 1200 ms[1]—in keeping with the intervals utilized in other paradigms involving complex visual stimuli (West & Holcomb, 2002). Measurements were subjected to repeated-measures ANOVA with the factors of Gesture Congruency (Congruous or Incongruous) and Electrode Site (29 levels). Because the relationship between the cartoon and the gestures was more obvious for some stimuli than for others, the ERPs were further subdivided according to each participant's median response latency into early and late decision trials. That is, each participant's ERPs to congruous and incongruous gestures were divided into early and late categories based on a median split of reaction times in the congruency task. A second repeated-measures ANOVA was performed with the additional factor of Decision Time (Early, Late).

ERPs elicited by pilot probe words were averaged over 1-s intervals. On the basis of visual inspection of the data, the mean

---

[1] In response to a reviewer's query, the time course of congruency effects was assessed by performing repeated-measures ANOVAs on the mean amplitude of ERPs within consecutive 50-ms time windows from 100 to 1200 ms. Statistically reliable effects of gesture congruency were observed continuously from 300 to 1200 ms.

amplitude of averaged waveforms was measured from 300 to 500 ms (N400) and from 500 to 900 ms. Measurements underwent a 2×2 repeated-measures ANOVA with the factors of Word Relatedness (Related or Unrelated) and Gesture Congruency (Congruous or Incongruous), along with Electrode Site (29 levels).

For all analyses, original degrees of freedom are reported; however, where appropriate, p values were subjected to Geisser–Greenhouse correction (Geisser & Greenhouse, 1959).

Results

Accuracy

Participants correctly classified 81% (*SE* 2%) of congruous gestures and 91% (*SE* 2%) of incongruous ones. Thus, participants related semantic information in gestures to that in the cartoons at a rate well above chance. A comparison between mean accuracy rates revealed that incongruous gestures were categorized more accurately, $F(1,11) = 9.9, p < .001$ (though due to artifact rejection, the mean number of trials in each condition was roughly equal).

Response Latencies

For responses occurring within 3 s after stimulus onset, participants classified congruous gestures (1345 ms, *SE* 68) reliably more quickly than incongruous ones (1510, *SE* 62), $F_1(1,11) = 9$, $p < .05$. The congruity effect was also reliable with items as the random variable, $F_2(1,159) = 17, p < .0001$. To confirm that this effect was not an artifact of excessive trimming, an additional analysis was conducted on responses within a 4-s window, which also proved robust, $F_1(1,11) = 8.9, p = .01$ (congruous: 1952 ms, $SE = 76$; incongruous: 2148 ms, $SE = 75$).

ERPs

A large, broadly distributed negative onset potential can be observed, peaking around 225 ms. Congruency effects are apparent around 300 ms in the form of broadly distributed negative component peaking around 458 ms (N450) in response to both congruous and incongruous gestures, with more negative ERPs in the case of incongruous items (see Figure 2). Overlap with the larger onset negativity may have caused the early portion of the N450 effect to be less discernible at anterior locations. A positive-going deflection (LPC) peaking around 740 ms was also elicited by congruous items, resulting in an extended congruity effect evident until the end of the epoch.

The N300 and N450 components were assessed by measuring the mean amplitude of ERPs elicited between 300–400 ms and 400–600 ms after stimulus, respectively. Between 300 and 400 ms, incongruous gestures elicited more negative ERPs than congruous ones across the scalp (Congruency main effect: $F[1,11] = 23.4, p < .0005$). Between 400 and 600 ms, a main effect of Congruency was also obtained, $F(1,11) = 100.0$, $p < .0001$, qualified by an interaction with electrode site, $F(28,308) = 7.0, p < .001, \varepsilon = .11$. During this time window, the congruency effect was largest over frontal and fronto-central midline scalp sites (Fz and FCz), due to more negative ERPs elicited by incongruent items in this region. (See Table 2 to compare mean amplitudes elicited over midline electrode sites.)

Incongruous gestures continued to result in greater negativity between 600 and 900 ms (Congruency main effect: $F[1,11] = 41.6$, $p < .0001$; Congruency × Electrodes interaction: $F[28,308] = 5.5, p < .005, \varepsilon = 0.125$), and 900 to 1200 ms poststimulus (Congruency main effect: $F[1,11] = 27.0, p < .0005$; Congruency × Electrodes interaction: $F[28,308] = 4.6, p < .01, \varepsilon = 0.1$), again with maximal effects at the fronto-central midline.

Median Split: Early versus Late Decision Trials

RTs. Response latencies were reliably shorter for congruous (1353 ms) than incongruous (1542 ms) gestures, $F(1,11) = 8.5$, $p < .05$. Congruency effects were approximately the same size in early (199 ms) and late decision (167 ms) trials, as no interaction between Congruency and Decision Time was observed, $F(1,11) = 1.3$, n.s.

ERPs. As shown in Figure 3, between 300 and 400 ms, congruency effects began earlier in early decision trials, where the relationship to the preceding cartoon context was apprehended more readily, than in congruent late decision trials (Congruency × Decision Time interaction 300–400 ms: $F[1,11] = 8.4, p < .05$). Follow-up analyses revealed main effects of congruency for early decision trials, $F(1,11) = 32.0$, $p < .0005$, but not late decision ones, $F < 1$, n.s. Similarly, between 400 and 600 ms, congruency effects were found only in early decision ERPs, due to the fact that congruent early decision trials elicited more positive ERPs than their late decision counterparts (Congruency × Decision Time interaction: $F[1,11] = 71.3$, $p < .0001$; Congruency main effect in early decision trials: $F[1,11] = 150.0$, $p < .0001$; Congruency main effect in late decision trials: $F[1,11] = 3.0$, n.s.). However, between 600 and 900 ms, main effects of Congruency were obtained at both levels of Decision Time, though the effect was larger and more robust for early decision trials (Congruency × Decision Time interaction: $F[1,11] = 15.0$, $p < .005$; Congruency main effect in early decision trials: $F[1,11] = 50.0$, $p < .0001$; Congruency main effect in late decision trials: $F[1,11] = 4.4, p = .06$). Between 900 and 1200 ms, the size of the congruency effect was similar for both levels of Decision Time (Congruency main effect: $F[1,11] = 28.0$, $p < .0005$; Congruency × Decision Time interaction: $F < 1$, n.s.).

Pilot Probe Words

All probe words elicited a broadly distributed N1/P2 complex followed by an N400. Unrelated words elicited more negative ERPs from approximately 300 ms poststimulus to the end of the epoch (900 ms). Within the time window in which the N400 is typically observed (300 and 500 ms), the mean amplitude of ERPs was shown to be reliably more negative for unrelated words relative to related ones (Relatedness main effect: $F[1,11] = 37.0$, $p < .0001$). Between 500 and 900 ms poststimulus, unrelated words continued to elicit more negative ERPs (Relatedness main effect: $F[1,11] = 31.0, p < .005$; Relatedness × Electrode Site interaction: $F[28,308] = 4.0, p < .05, \varepsilon = .11$). These findings demonstrate that probe words elicited intended N400 relatedness effects and were suitable for use in Experiment 2.

Discussion

At least two distinct ERP components contributed to the observed gesture congruity effect. Both congruous and incongruous gestures elicited a negative-going deflection peaking approximately 450 ms poststimulus (N450), though N450 amplitude was much greater for incongruous items. Congruous gestures also evoked a broadly distributed positivity peaking around 740 ms (LPC).

The N450 observed in the present study to contextually incongruous gestures is similar to the N400 observed in ERP stud-

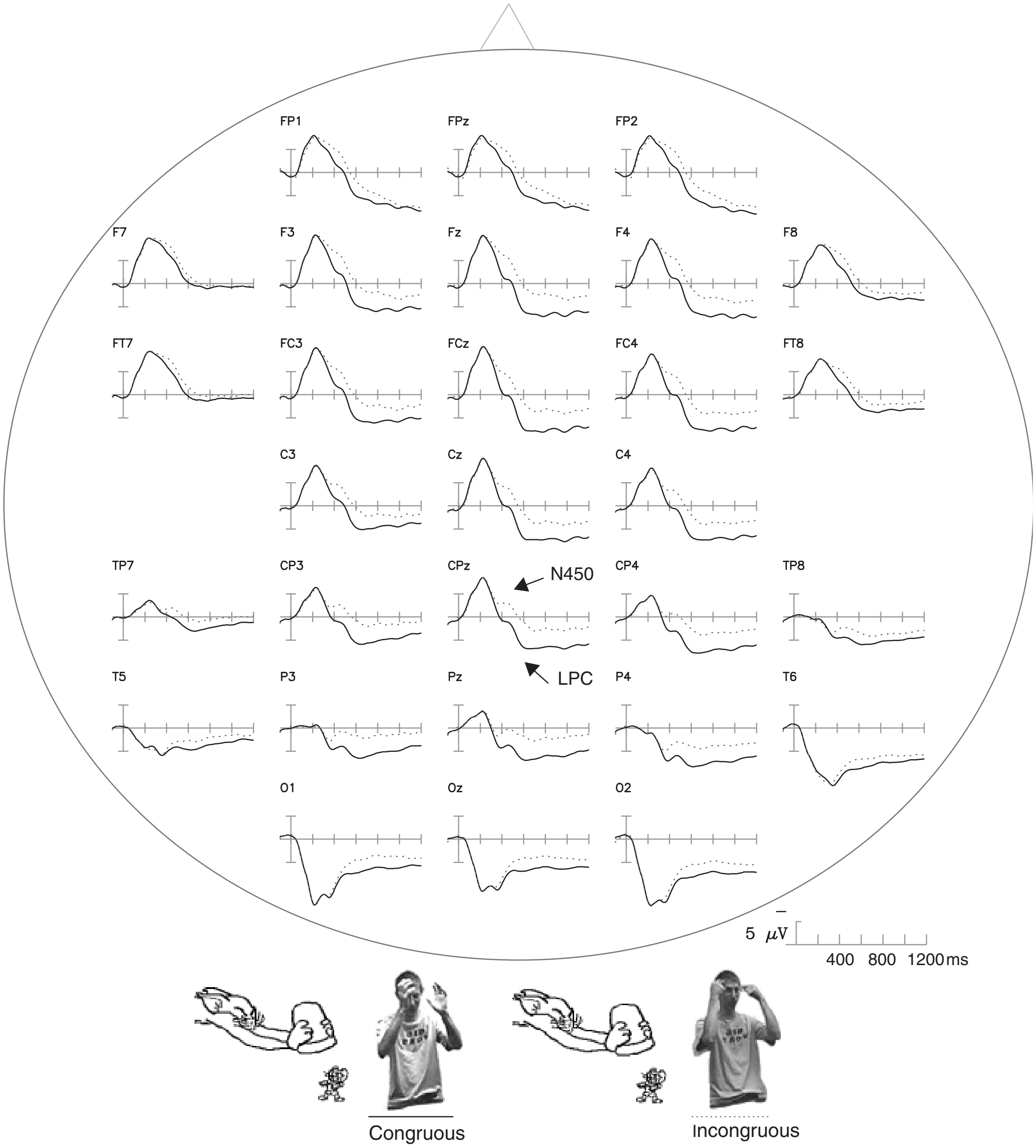

Figure 2. Experiment 1: ERP responses time-locked to the onset of congruous and incongruous gestures and extending for 1200 ms. Time zero corresponds with the onset of gesture clips.

ies of image-based tasks involving line drawings, photographs, picture stories, and videos. For example, Barrett and Rugg (1990) asked subjects to make relatedness judgments for pairs of sequentially presented pictures and observed a larger N450 for the second picture in an unrelated (wrench–fork) than a related (knife–fork) pair. As in the present study, most such ''picture'' ERP studies report a broadly distributed negativity largest at frontal electrode sites and not evident at occipital sites (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999; Sitnikova, Kuperberg, & Holcomb, 2003; West & Holcomb, 2002).

Although incongruous gestures elicited more negative ERPs even earlier, between 300 and 400 ms poststimulus, we suggest that this early effect reflects the onset of the N450 rather than the N300. Unlike the anterior focus typical of the N300, the congruity effect that we observed 300–400 ms poststimulus was broadly distributed over the scalp. In this respect, our findings are similar to those of Sitnikova and colleagues (2003), in whose study participants viewed videotaped action sequences with appropriate and inappropriate objects (e.g., shaving with a razor vs. a rolling pin). These researchers report an N400-like response to inappropriate objects with an onset around 325 ms poststimulus, but no discernible N300 (Sitnikova et al., 2003). The absence of a distinguishable N300 in the present paradigm may reflect true differences in the processing of static as opposed to moving images, or it may simply be an artifact of the rapid presentation parameters necessary for videographic stimuli. That is, the variable onset of stimulus recognition might preclude consistent time-locking to the neural activity necessary to elicit an N300 distinct from subsequent N400-like activity.

Table 2. *Experiment 1: Mean Amplitude and Standard Error (in Microvolts) of ERPs Recorded over Fz, Cz, Pz, and Oz*

	Electrode site			
Interval (in ms)	Fz	Cz	Pz	Oz
300–400				
Congruent	− 7.2 ± 1.0	− 6.3 ± 1.0	0.5 ± 1.5	9.6 ± 1.3
Incongruent	− 9.2 ± 1.2	− 8.0 ± 1.0	− 1.0 ± 1.5	8.6 ± 1.5
400–600				
Congruent	− 0.6 ± 1.5	0.4 ± 1.4	2.6 ± 1.2	6.4 ± 1.3
Incongruent	− 5.4 ± 1.6	− 3.8 ± 1.4	− 1.1 ± 1.3	4.2 ± 1.2
600–900				
Congruent	4.8 ± 1.6	4.0 ± 1.1	4.2 ± 1.1	5.3 ± 1.0
Incongruent	0.2 ± 1.7	0.0 ± 1.2	0.2 ± 1.0	2.6 ± 1.0
900–1200				
Congruent	4.6 ± 1.4	3.3 ± 1.0	3.8 ± 1.0	5.2 ± 1.0
Incongruent	1.1 ± 1.5	− 0.4 ± 0.8	0.1 ± 1.0	2.9 ± 1.3

The peak latency of the gesture N450 component is consistent with the time course of the N400-like effect obtained in response to action video clips (Sitnikova et al., 2003) and picture stories (West & Holcomb, 2002). In contrast, the N400 elicited by pictures of individual objects tends to peak slightly earlier. This pattern of outcomes suggests that comprehending gestures—like the comprehension of actions (Sitnikova et al., 2003) and illustrations of detailed scenes (West & Holcomb, 2002)—takes longer than the processing of static images of single objects. One potential contribution to the increased processing load is the dynamicity and visual complexity of the gesture clips that we used as stimuli. In keeping with this proposal, analogous N400 peak latency shifts have been observed in paradigms designed to tax perceptual processes, such as auditory masking (Connolly, Phillips, Stewart, & Brake, 1992) or visual stimulus degradation (Holcomb, 1993).

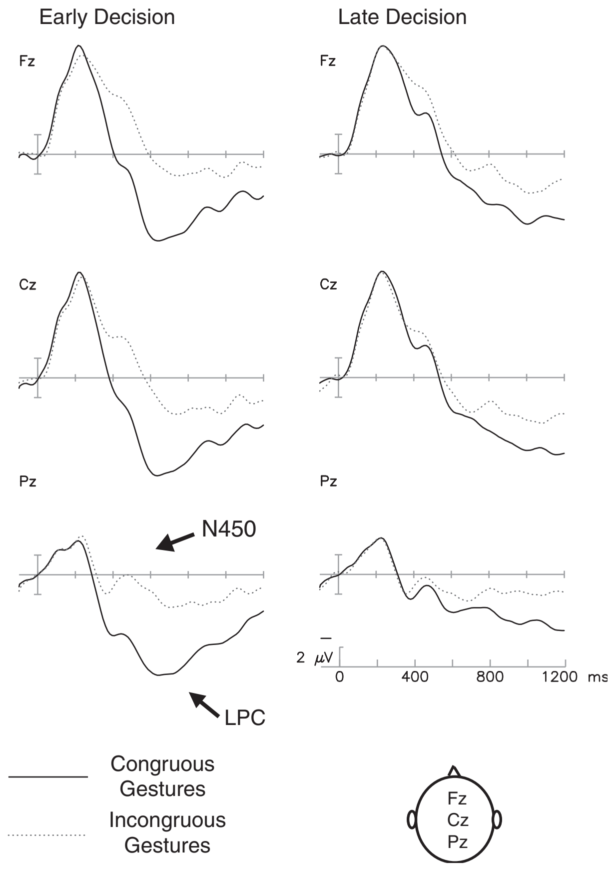

Figure 3. Experiment 1: Gesture ERPs sorted according to subjects' median decision times. The N450 effect occurs primarily in cases where the relationship between cartoon and gesture was readily apparent. Congruency effects occur later for late decision trials. Time zero corresponds with the onset of gesture clips.

Overall, the time course, morphology, and functional characterization of the gesture N450 suggests that it indexes a semantic integration process similar to that underlying the picture-priming N400 and analogous to that underlying the classic N400 elicited by verbal stimuli. Reaction time data paralleled this outcome: Incongruous gestures took (on average) 165 ms longer to classify, as would be expected if participants were attempting to integrate salient elements of gestures and cartoons. The interpretation of the gesture N450 as an index of semantic integration is further supported by the finding that the N450 effect (i.e., measured 400–600 ms after gesture onset) was evident only in early decision trials, where participants were able to rapidly apprehend congruency relations.

One must exercise caution, however, in attributing the observed N450 effect exclusively to neural activity associated with semantic integration, as this effect was obviously driven in part by overlap with the positivity to congruous items (peaking approximately 740 ms after stimulus onset). That this positivity is enhanced to congruent items suggests its membership within the P300 family of potentials, which reflect brain activity associated with stimulus evaluation and categorization, and which are larger in amplitude in response to targets (for review, see Johnson, 1988; Kok, 2001; Pritchard, 1981; Soltani & Knight, 2000). P300 is often associated with binary decision tasks, and thus may have been engendered by the gesture classification task used in the present study. In such paradigms, P300 latency is typically correlated with RTs on the decision task. Accordingly, in the present study, the LPC to congruous items peaked earlier in early decision trials, than in late decision trials.

EXPERIMENT 2

In Experiment 2, we sought to dissociate the overlapping contributions of the N450 and the LPC by utilizing a paradigm that did not require an overt response to gestures. The stimuli and procedures from Experiment 1 were repeated. However, participants were asked to classify the related and unrelated probe words that followed each cartoon/gesture pair rather than the gestures themselves. If the observed effect of gesture congruency was driven exclusively by the positive-going LPC elicited by the congruent gestures, we would expect to see no difference in the amplitude of the negative component observed between 400 and 600 ms. On the other hand, if congruency effects in Experiment 1 reflected an N400-like component, then the gestures in Experiment 2 should also elicit more negative ERPs when they are contextually incongruous than when they are congruous.

Method

Participants

Sixteen healthy, fluent speakers of English with no history of neurological impairment were recruited for the study. None had participated in Experiment 1 or any of the normative experiments associated with the study. The data from 4 participants were excluded due to excessive eye movements and other artifacts. Of the remaining 12 participants, 10 were right-handed and 2 were left-handed. The mean laterality quotient was .615 for right-handers and − .5 for left-handers.

Materials and Procedure

Stimuli and procedures were identical to those used in Experiment 1 with the exception that participants were instructed to read the probe word that appeared on the screen and press the YES button if they felt confident that the word related to some element of the preceding context or the NO button if the word was unrelated. Typically, related words denoted objects or actions depicted in both the cartoon and the congruous gestures, and would elicit a YES response if preceded by either a related cartoon and congruous gesture or a related cartoon and incongruous gesture.

As in Experiment 1, a 2 × 2 within-subjects Congruency × Relatedness design was employed. Each participant saw 80 congruous and 80 incongruous cartoon/gesture pairs, as well as 80 related and 80 unrelated probes. Four lists were constructed such that no word or video clip was repeated on any list, but across lists, each word appeared once as a related item and once as an unrelated item following either a congruous or incongruous video context. In subsequent discussion, "congruency" will refer to the relationship between the preceding cartoon and gesture, and "relatedness" to the relationship between the probe word and the preceding context.

Normative Study

Related and unrelated probe words were selected on the basis of experimenters' intuitions. To verify semantic correspondences between probes and videos, 40 volunteers from the University of California, San Diego, community viewed each trial and classified probes as either related or unrelated to their preceding cartoon/gesture contexts. Related words following congruous cartoon/gesture pairs were correctly classified on 89% ($SE = 2\%$) of items; related words following incongruous pairs were correctly classified on 85% ($SE = 2\%$) of items. Unrelated words following both congruous and incongruous pairs were classified accurately on 96% ($SE = 1\%$) of trials. These data indicate that most of the intended relations between words and their preceding contexts were consistently recognizable.

Data Analysis

Behavioral data were analyzed in a manner identical to Experiment 1. Analyses by subjects were conducted on responses occurring within a 3.8-s window after stimulus presentation. Only 0.3% trials total were lost due to trimming.

EEG Recording

Data recording and the construction of ERPs proceeded as in Experiment 1. On average, there were 50 trials in each critical gesture bin, and the artifact rejection rate was 29% in both cases. For probe words, there were an average of 29 trials in critical bins ($SD = 5$), and an average artifact rejection rate of 20% ($SD = 12\%$).

Gesture Analysis

The mean amplitude and peak latencies of the waveforms were measured within the same four time windows utilized in Experiment 1: 300–400, 400–600, 600–900, and 900–1200[2] ms after onset. All measurements were subjected to repeated-measures ANOVA with the factors of Gesture Congruency (congruous or incongruous) and Electrode Site (29 levels).

To assess differences in congruency effects elicited by gestures whose meaning was more or less readily apparent, trials from the present experiment were sorted according to median response latencies from Experiment 1, as well as the congruency ratings obtained through the normative study described in Experiment 1. Congruous trials classified in Experiment 1 before the median response latency were binned as highly congruent items, with a mean congruency rating of 4.5, $SD = 0.57$ (incongruent: 1.3, $SD = 0.37$); remaining trials were binned as moderately congruent items, with a mean rating of 3.8, $SD = 0.83$ (incongruent: 1.5, $SD = 0.43$). A second repeated-measures ANOVA was conducted on the new alignment of data with the additional factor of Rating Level.

Probe Word Analysis

Measurements were analyzed with repeated-measures ANOVA within the same time intervals as Experiment 1 (300–500 ms and 500–900 ms) with the factors of Word Relatedness (related or unrelated), Gesture Congruency (congruous or incongruous), and Electrode Site. The application of Geisser–Greenhouse correction for both probe word and gesture data were conducted in the same manner as in Experiment 1.

Results

Gestures

Both congruous and incongruous items elicited a large negative onset potential, peaking around 230 ms (see Figure 4). The distribution of this component is similar to that observed in Experiment 1: The amplitude of the negativity decreases over posterior sites (Pz, P3, P4) and inverts in polarity over occipital sites (Oz, O1, O2). The onset negativity was followed by a second broadly distributed negativity (N450) peaking around 462 ms for congruous items and 476 ms for incongruous ones. In contrast to Experiment 1, no effects of congruency were obtained between 300 and 400 ms, all Fs <1, n.s.

Measured between 400 and 600 ms, the N450 was reliably larger for incongruous than congruous gestures (Congruency main effect: $F[1,11] = 12.4$, $p < .005$). Congruency effects were also obtained in the 600–900-ms window (main effect: $F[1,11] = 29.0$, $p < .0005$; Congruency × Electrode Site interaction: $F[28,308] = 4.0$, $p < .01$, $\varepsilon = .14$), and the 900–1200-ms window (main effect: $F[1,11] = 22.0$, $p < .005$; Congruency × Electrode Site interaction: $F[28,308] = 3.3$, $p < .05$, $\varepsilon = .15$). As shown in Table 3, the congruity effect was largest over Cz and parietal sites in all time windows. Unlike Experiment 1, positive-going, LPC-like deflections of the waveform were not observed in this study.

Median Split: Highly versus Moderately Congruent Trials

Gesture congruency effects in highly and moderately related trials can be seen in Figure 5. Measured from 300 to 400 ms, no effects were observed, all Fs <1, n.s. Between 400 and 600 ms, ERPs to incongruous gestures were more negative than congruous ones, $F(1,11) = 15.0$, $p < .005$. This congruency effect appears larger in the case of highly congruent trials. Although the interaction between Gesture Congruency and Rating Level only approached significance, $F(1,11) = 4.0$, $p = .07$, follow-up contrasts revealed that the congruency effect in highly

---

[2] Again, repeated-measures ANOVAs were performed on the mean amplitude of ERPs within consecutive 50-ms time windows from 100 to 1200 ms. Statistically reliable effects of gesture congruency were observed continuously from 450 to 1200 ms.

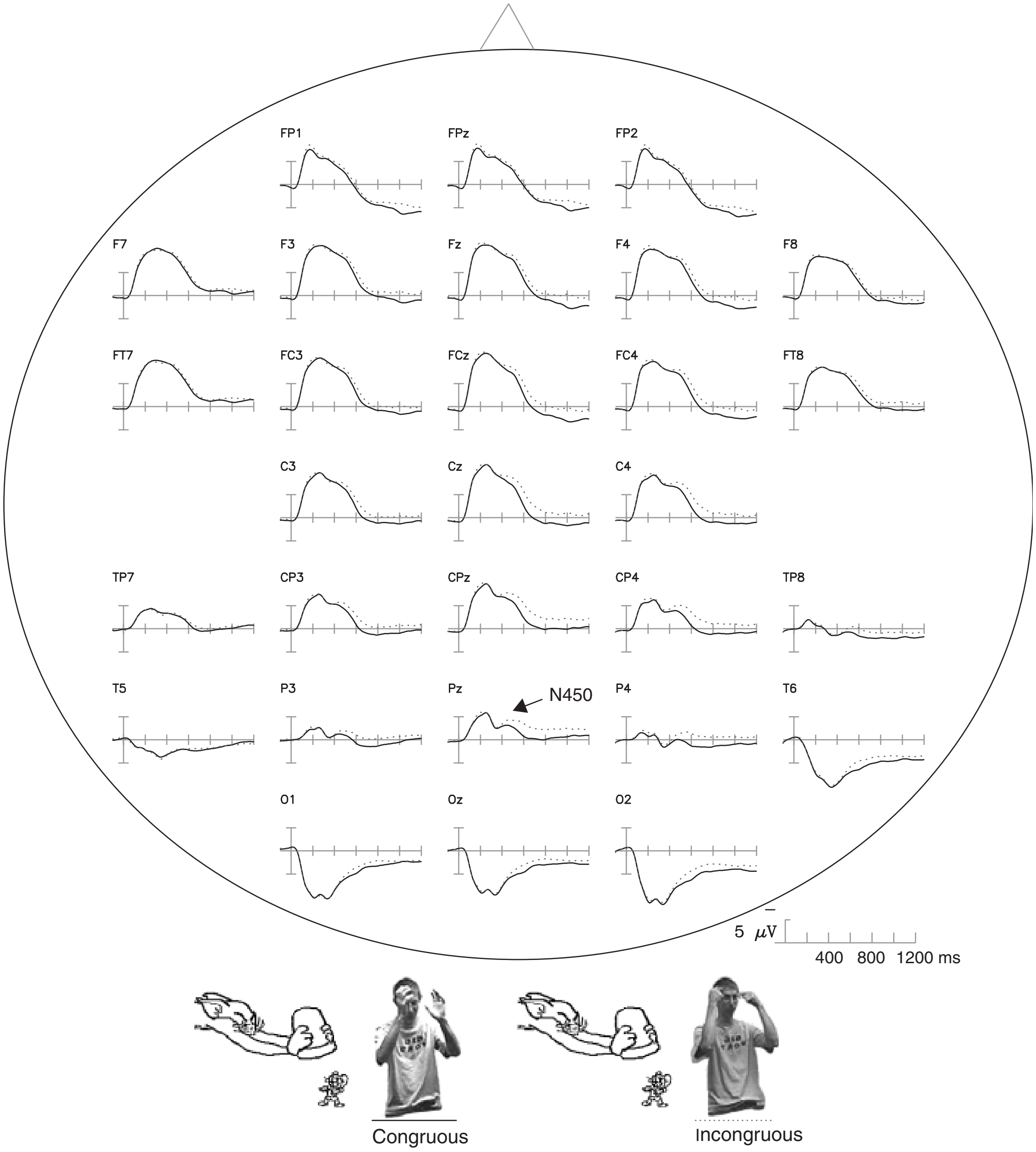

Figure 4. Experiment 2: Grand averaged ERPs time-locked to the onset of congruous and incongruous gestures and extending for 1200 ms. Time zero corresponds with the onset of gesture clips.

congruent trials was robust, $F(1,11) = 9.0$, $p = .01$, whereas in moderately congruent trials, neither the main effect of Congruency nor the interaction with Electrode Site proved reliable, Fs < 1, n.s.

From 600 ms poststimulus to the end of the epoch (1200 ms), both types of incongruent trials continued to elicit more negative ERPs than congruent ones (Congruency main effect: 600–900, $F[1,11] = 10.0, p < .01$; 900–1200, $F[1,11] = 30.0, p < .0005$). However, within the 900–1200-ms time window, the main effect of Congruency was qualified by a two-way interaction with Electrode Site, $F(28,308) = 2.8$, $p < .05$, $\varepsilon = .13$, and a marginally significant three-way interaction with Rating Level and Electrode Site, $F(28,308) = 2.5$, $p = .07$, $\varepsilon = .11$. The three-way interaction reflects the attenuation of the congruency effect over frontal electrode sites for highly, but not moderately related trials (see Figure 5).

Word Response Latencies

A main effect of Relatedness was found with both subjects, $F_1(1,11) = 5$, $p < .05$, and items, $F_2(1,159) = 32$, $p < .0001$, as random variables, revealing faster responses to related items. A main effect of Congruency, $F_1(1,11) = 6.5$, $p < .05$; $F_2(1,159) = 19.3$, $p < .0001$, was also observed, as responses were faster to congruous than incongruous items (see Table 4). A reliable interaction between Congruency and Relatedness was observed in the analysis by trials, $F_2(1,159) = 6.5, p < .05$, but not by subjects, $F_1(1,11) = 2.5$, $p = .14$. Follow-up analyses revealed that related words following congruous cartoon/gesture pairs were classified faster than the same words following incongruous pairs, $F_1(1,159) = 19.3$, $p < .0001$, whereas the effect of gesture congruency on responses to unrelated words did not reach significance, $F_1(1,159) = 2.8, p = .1$. This pattern of outcomes suggests

Table 3. *Experiment 2: Mean Amplitude and Standard Error (in Microvolts) of ERPs Recorded over Fz, Cz, Pz, and Oz*

	Electrode Site			
Interval (in ms)	Fz	Cz	Pz	Oz
400–600				
Congruent	−8.4 ± 1.3	−7.9 ± .94	−3.6 ± .64	5.0 ± 1.0
Incongruent	−9.5 ± 1.2	−9.8 ± .89	−5.0 ± .78	4.3 ± 1.6
600–900				
Congruent	−1.3 ± 1.2	−1.2 ± .91	−1.5 ± .44	2.5 ± 0.8
Incongruent	−2.9 ± 1.3	−3.8 ± .81	−3.5 ± .57	1.6 ± 1.0
900–1200				
Congruent	1.0 ± 0.9	−0.1 ± .62	−2.0 ± .50	5.2 ± 1.0
Incongruent	−1.3 ± 1.0	−2.3 ± .70	−3.4 ± .72	2.9 ± 1.3

that participants benefited from congruous gestures when classifying related words, but less so for unrelated ones. This finding is bolstered by evidence that participants were more accurate in classifying words following congruous gestures.

Word Accuracy

Probes were classified slightly more accurately when following congruous cartoon/gesture pairs than incongruous ones, $F(1,11) = 4.4$, $p < .05$. However, no main effect of Word Relatedness or interaction was observed, $F < 1$, n.s. (see Table 4).

Word ERPs

Early activity elicited by related words in both congruous and incongruous conditions includes a broadly distributed N1 component peaking at ~ 100 ms and a P2 component peaking at ~ 190 ms. Differences in brain responses are first observable between 350 and 420 ms: Related probe words preceded by incongruous gestures evoked a broadly distributed negative waveform (N400) peaking around 367 ms, whereas probe words preceded by congruous gestures resulted in a smaller negativity that peaked around 353 ms. After 500 ms, both probe word types were associated with extended positive-going activity peaking around 725 ms and continuing to the end of the epoch (900 ms) (see Figure 6).

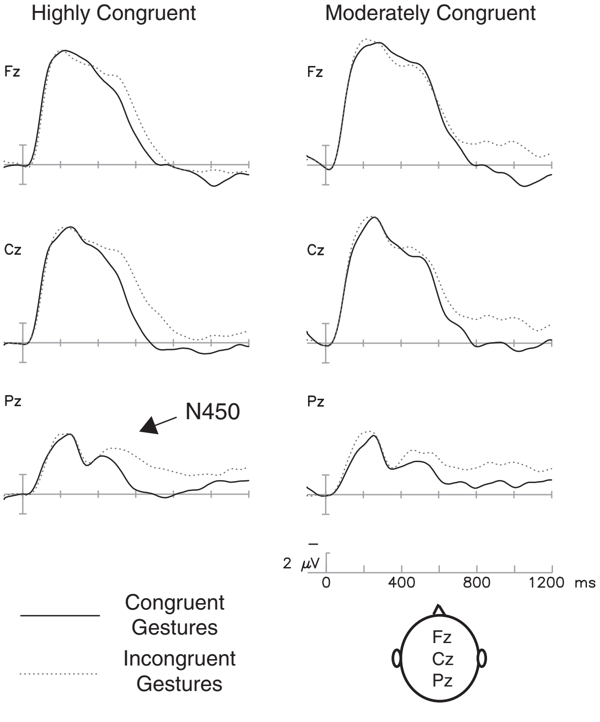

Figure 5. Experiment 2: Grand averaged ERPs time-locked to the onset of highly and moderately congruent gestures. N450 effects are larger for highly related items; late congruency effects are more prominent for moderately related ones.

Table 4. *Mean Accuracy and Response Latencies in Classifying Probe Words*

	Congruent		Incongruent	
	Mean accuracy	Mean RT (ms)	Mean accuracy	Mean RT (ms)
Related	.95 (.01 *SE*)	1114 (29 *SE*)	.89 (.02 *SE*)	1303 (35 *SE*)
Unrelated	.94 (.02 *SE*)	1338 (28 *SE*)	.91 (.03 *SE*)	1392 (27 *SE*)

Between 300 and 500 ms poststimulus, main effects of both Word Relatedness, $F(1,11) = 24.0$, $p < .0005$, and Gesture Congruency, $F(1,11) = 5.0$, $p = .05$, were obtained, qualified by an interaction between these two factors, $F(1,11) = 6.7$, $p < .05$. Follow-up analyses demonstrated a robust effect of Gesture Congruency on the amplitude of the N400 elicited by related words, $F(1,11) = 19.1$, $p < .005$. However, in the case of unrelated words, neither the main effect of Gesture Congruency nor the interaction with Electrode Site approached significance, Fs < 1.5, n.s.

Between 500 and 900 ms, unrelated words continued to elicit more negative ERPs than related ones (main effect of Word Relatedness, $F[1,11] = 13.3$, $p < .05$), but neither the main effect of Gesture Congruency nor the interaction between Word Relatedness and Gesture Congruency reached significance.

Discussion

Experiment 2 yielded two main findings. First, the N450 effect of gesture congruency on ERPs was replicated in a procedure that did not demand participants' explicit classification of gestures. Beginning approximately 400 ms poststimulus, the presentation of incongruous gestures again resulted in more negative ERPs in comparison with congruous trials. Furthermore, in this case, congruous gestures did not additionally elicit a positive-going component within the same time window, suggesting a dissociation in the cognitive processes giving rise to the N450 and LPC in Experiment 1. These findings support our suggestion that the LPC elicited by gestures in Experiment 1 was task driven. The N450, on the other hand, appears to be driven by processes sensitive to the congruency relations between a gesture and its preceding context. This pattern of outcomes corroborates a view of the N450 observed in Experiments 1 and 2 as being analogous to the N400 component elicited by interpretable pictures and words.

Further support for this idea is advanced by the finding that the amplitude of the N450 was modulated by the degree of congruency between gestures and cartoons. Gesture N450 effects were larger and more robust for highly than moderately congruent trials. In contrast, between 600 and 900 ms after stimulus onset, comparable congruency effects were observed for both types of trials. Between 900 and 1200 ms, these effects were

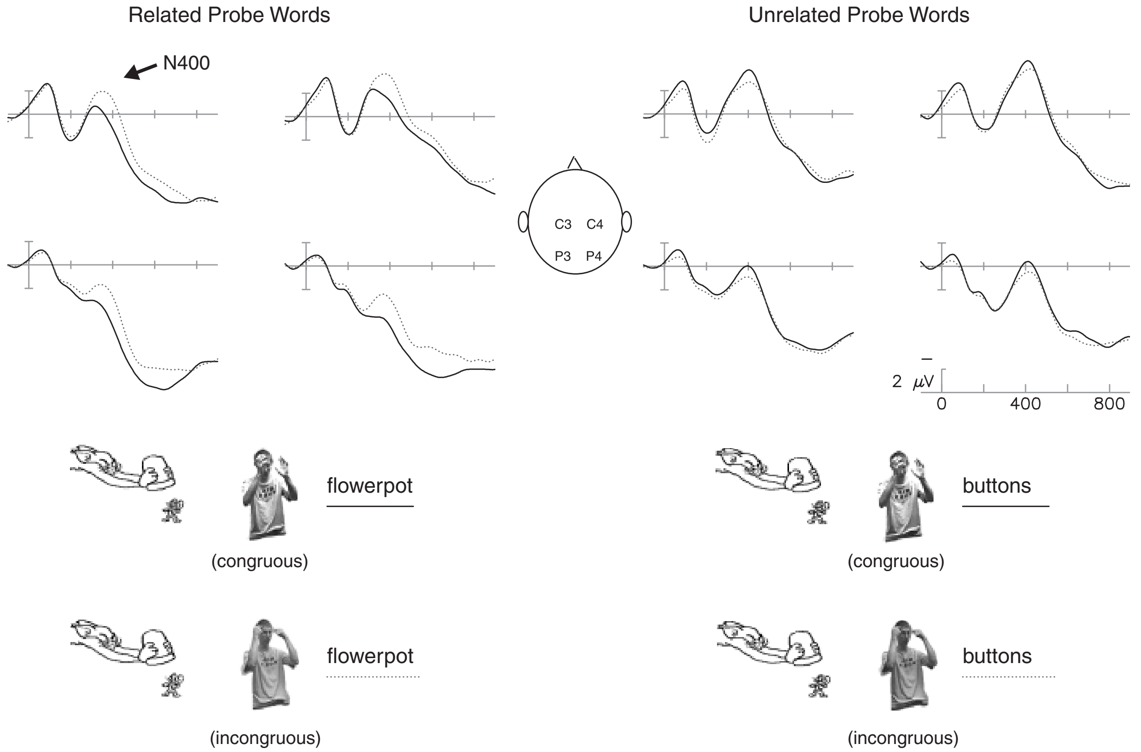

Figure 6. Experiment 2: N400 elicited by related and unrelated words following congruous and incongruous contexts.

greatly reduced for highly but not moderately congruent items. These latter results suggest that the semantic processing of moderately congruent gestures, which are more difficult to integrate with context, starts later and continues longer relative to the highly congruent gesture trials.

One notable difference from Experiment 1 is the absence of an early effect of gesture congruency (between 300 and 400 ms poststimulus). Early congruency effects in Experiment 1 were likely due to the gesture classification task. In Experiment 2, in contrast, gesture congruency was not directly task relevant. Thus, it is all the more remarkable that the ERPs elicited by congruous and incongruous gestures in Experiments 1 and 2 exhibited general similarities, though the gesture N450 effect was smaller in Experiment 2—presumably due to the absence of the overlapping positivity elicited in Experiment 1.

The second major finding of this study was the modulation of word comprehension by congruous gestures. Related probes following congruous and incongruous gestures were identical, and both required a YES response. Nevertheless, the N400 was reliably larger for related words following incongruous gestures. No such effect of gesture congruency was observed on the amplitude of ERPs elicited by unrelated words (see Figure 6). An analogous pattern of results was observed in response latencies for classifying words.

This advantage for processing words preceded by congruous gestures relative to incongruous ones is consistent with results reported in other lexical priming studies using primes designed to activate perceptual features shared by the meaning of the target. For example, names of concrete objects (*hat*, *door*) resulted in attenuated N400 when preceded by corresponding object pictures relative to pictures of different objects (Pratarelli, 1994). Conversely, the same pictures resulted in reduced N400, but with a more anterior focus, when preceded by corresponding object names. A number of reaction time studies have demonstrated similar cross-modal priming of words and pictures (Carr, Sperber, McCauley, & Parmalee, 1982; Coney & Abernathy, 1994; Hines, 1993; Pratarelli, 1994; Vanderwart, 1984).

Within this framework, the present study demonstrates that even the highly schematic, evanescent visuo-spatial and motoric information in gestures is sufficient to affect the processing of related words. This view is consistent with the finding reported by Kelly et al. (2004) that ERPs time-locked to the auditory presentation of words are modulated by gesture congruency both during early auditory and subsequent semantic analysis. Presumably, words related to objects and features just activated in memory by gestures would be easier to process than unrelated lexical items. However, because a similar advantage for related words preceded by congruous gestures was not observed in Experiment 1, it is possible that explicit attention to the semantic relationship between a word and its context is necessary for gestures to modulate the comprehension of words.

General Discussion

Gesture N450

Event-related potentials were used to explore a number of questions related to gesture comprehension. In Experiment 1, manipulations of the congruency relationship between iconic gestures and their preceding context resulted in an enhanced negative component peaking around 450 ms poststimulus (N450) for incongruous trials. A similar effect was observed in Experiment 2, which utilized the same materials but did not require overt classification of gestures. These outcomes suggest that the congruency effects in both experiments were driven to some extent by genuine differences in semantic processing of congruous and incongruous gestures.

We suggest that the gesture N450 observed in the present study is a member of the N400 class of negativities, which are responsive to manipulations of relatedness and semantic constraint across a range of modalities and experimental paradigms. For example, high cloze, or preferred, sentence endings tend to elicit attenuated N400 amplitude, whereas low cloze, or unlikely, endings elicit large ones (Kutas & Hillyard, 1984). Likewise, when participants are presented with sequences of sentences (van Berkum et al., 1999) or pictures (West & Holcomb, 2002) representing a series of events, compatible successors result in reduced N400 relative to incompatible ones. In fact, N400-like responses have also been elicited by videos of actions (such as shaving performed with appropriate or inappropriate [razor vs. rolling pin] objects; Sitnikova et al., 2003). In all of these cases, coherent pictorial, lexical, and videographic sequences serve to activate stored knowledge in LTM, prompting expectations within the comprehender about the semantic content of upcoming information. Ensuing N400 effects can be viewed as an index of how well the stimulus matches expectations engendered by knowledge already active in the comprehender's working memory.

By analogy, in the present study, our cartoons presumably activated stored knowledge and engendered expectations about the meaning of the gestures that followed them. Insofar as iconic gestures are a semiotic medium parallel to words and pictures, gestures that prompted mental activity consistent with the observer's interpretation of the cartoon elicited reduced N450 relative to gestures that activate unexpected information. In this view, the gesture N450 reflects the semantic integration of gesture-based information into a higher order conceptual model. Further, although the more central distribution of the gesture N450 relative to the classic verbal N400 suggests that slightly different brain areas generate these components (see Figure 7), their shared sensitivity to contextual congruity may indicate membership in a family of interrelated neural processes subserving contextual integration in different modalities.

One concern raised by this paradigm is the possibility that verbal information attained through lipreading, rather than gestures themselves, was the source of the observed effects. We find this proposal unlikely for a number of reasons. First, the speaker's mouth subtended less than 0.25° of visual angle, minimizing the discernibility of lip movement and other nonauditory information deriving from the physical production of speech. Secondly, a recent study in our laboratory revealed that similar N450 effects are elicited by static gesture "freeze frames"— ⊖ which do not represent any dynamic speech production information—extracted from the video streams used as stimuli in the present experiments (Wu, 2005).

A second question is whether these findings generalize to instances of everyday language use. Although iconic gestures are subject to semantic processing within a laboratory setting, it is possible that in conversation, they may not carry the same impact due to the concurrent demands of speech processing or inattentiveness on the part of the listener. Further, in the present study, the use of image- rather than language-based contextual cues created somewhat different cognitive demands from those involved in comprehending authentic co-speech gestures. Rather than semantically analyzing gestures, could participants have simply matched perceptual features of the speaker's movements and hand configurations with features of preceding cartoons still active in visuo-spatial working memory?

A number of facts argue against this view. First, in approximately half of the trials, the speaker's gestures depicted features of objects, such as a dog's front legs, an opened box, and an extended jaw (see Figure 8). In all of these instances, the gestures bear no direct resemblance to their referents; rather, their meaning derives from higher order categorical correspondences. For example, in a trial where the speaker's gesture depicted a falling candle, his arm does not really resemble a candle. However, by bending at the elbow and extending his forearm and hand upward, he enacts the original vertical orientation of the candle in the cartoon. In other cases, the speaker uses his hands or index fingers to trace the outline of an object, as when he made

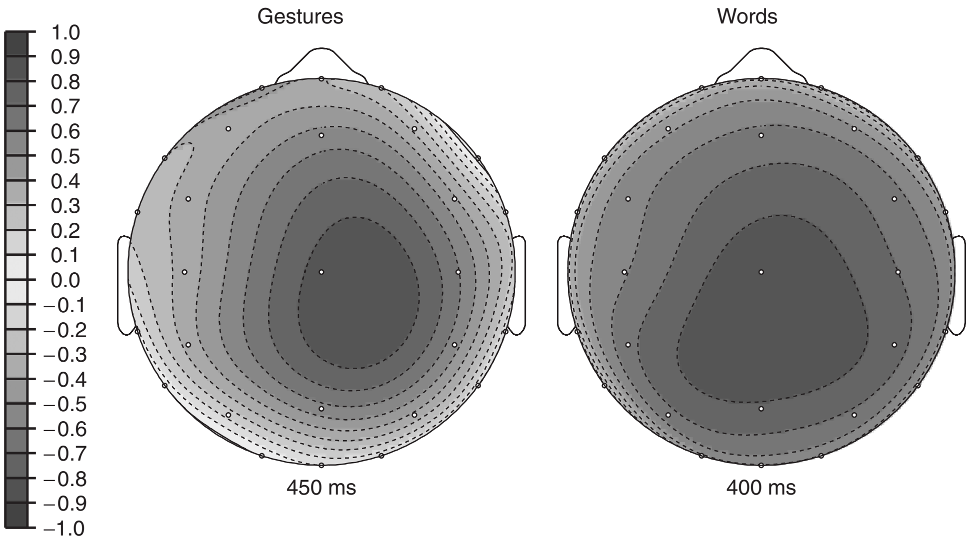

Figure 7. Experiment 2: Isovoltage maps show the distribution of gesture congruency effects at 450 ms postonset (incongruent minus congruent), and word relatedness effects at 400 ms postonset (unrelated minus related). Voltage in microvolts is represented on the scale bar. Values were normalized by dividing each data point for each stimulus type by the absolute value of the maximal data point for each stimulus type.

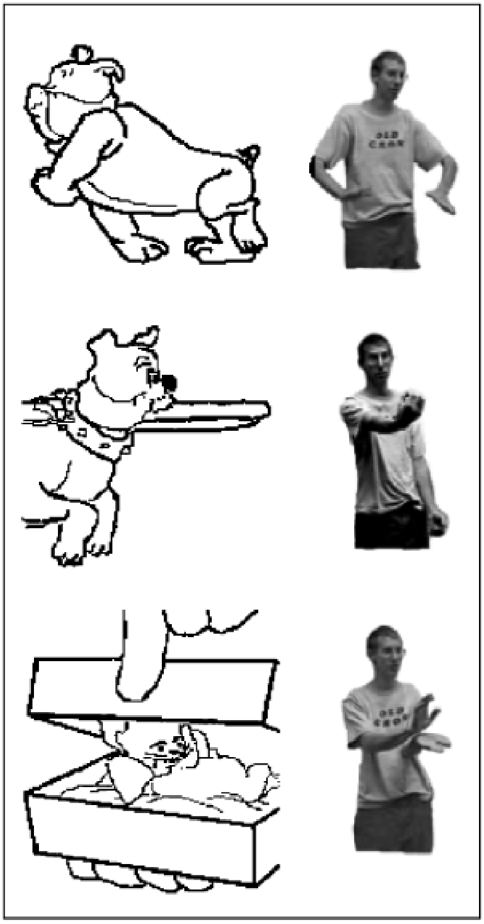

Figure 8. Examples of gestures used in Experiments 1 and 2. The speaker is portraying the dog's extended jaw, the opening box, and the dog's front legs.

iterative curving motions with extended index fingers to indicate the shape of a platter. Although these motions look nothing like the platter carried by the cat in the cartoon, they can nevertheless be mapped to an oval shape analogous to the shape of a platter.

In other cases, where the speaker's gestures depict actions performed in the cartoon, correspondences are still largely categorical rather than perceptual. Actions in the cartoons were often shown in profile, whereas the speaker faced the camera as he gestured (see Figures 5 and 8). In some cases, the speaker's gestures depict actions that were not actually shown in the cartoon, but which could be inferred. For example, in one trial, the speaker depicted how Elmer Fudd opens a large box with a crowbar-like instrument, though the cartoon clip only shows him lifting aside the lid. In other cases, the cartoon characters' actions involved rapid, wide-ranging motions (e.g., leaping, diving, flying), that differed from the gestural enactments. Finally, actions that occured on the left side of space in the cartoon were often transposed in the gestural depiction to the right side of space and vice versa.

Given these kinds of distinctions, apprehending relationships between gestures and cartoons is more likely to require conceptual integration (Coulson & Van Petten, 2002; Fauconnier & Turner, 1998, 2002) or "mesh" (Glenberg, 1997; Glenberg & Robertson, 1999) between perceptual input and stored knowledge about the phenomena being depicted, rather than direct matching. For example, in the flowerpot trial illustrated in Figure 1, the narrator's hands are held at an even height and maintain a C shape at a constant distance in a continuous, rapid arc while lowered from above the head to approximately chest level. This temporal and spatial coordination is consistent with what we know about the affordances and movement schemas involved in manipulating containers in order to trap small, elusive animals. Presumably, in the experiment, this kind of background knowledge was preactivated by cartoons, and the ensuing reduction in the amplitude of the N450 in response to congruous gestures indexed decreased resources devoted to mapping the visuo-spatial features of gestures to a mental representation of the event or object that they depict.

A useful analogy can be drawn in the domain of language. For example, comprehending figurative (and some instances of literal) language use is thought to require the apprehension of shared relational structure between distinct knowledge domains (Coulson & Van Petten, 2002). A metaphorical sentence such as, "After giving it some thought, I realized the new idea was a gem," invites the reader to treat properties of ideas and gems as analogical counterparts. Just as the clarity of a gem, for instance, allows for the passage of light, so the clarity of an idea allows for the transfer of new insight. By contrast, in the use of the literal statement, "The stone we saw in the natural history museum was a gem," it is argued that considerably less retrieval and alignment of conceptual structure is necessary for a reader to apprehend the linguistically cued mapping between "that stone we saw in the natural history museum" and the category "gem." Like metaphorical language use, the comprehension of iconic gestures is also proposed to recruit relational mappings—in this case, between visuo-spatial structure visible in real time and conceptual structure active within the listener—to a greater degree than scenes involving explicit actions and objects.

Although further research is necessary to determine if this kind of integration process is actually engaged by gestures in the course of natural conversation, the present study demonstrates that movements produced by the hands and body known as iconic gestures are at least amenable to semantic integration processes. Moreover, the schematic movements comprised by iconic gestures differ from explicit actions in that the integration of a gesture's semantic features with other contextually active information may involve conceptual mapping (Fauconnier, 1985) to a greater degree than the comprehension of actions and events depicted in the picture story and video experiments described previously.

REFERENCES

Alibali, M. W., Flevares, L., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, *89*, 183–193.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*, 201–212.

Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed in speech? An experimental investigation. *Semiotica*, *123*, 1–30.

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology*, *18*, 438–462.

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*, *93*, 179–192.

Carr, T. H., Sperber, R. D., McCauley, C., & Parmalee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 757–777.

Cassell, J., McNeill, D., & McCullough, K. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition*, *7*, 1–33.

Coney, J., & Abernathy, M. (1994). Picture-name priming in the cerebral hemispheres. *Brain and Language*, *47*, 300–313.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*, 1–18.

Coulson, S., & Van Petten, C. K. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*, 958–968.

Efron, D. (1972). *Gesture, Race, and Culture*. Paris: Mouton.

Fauconnier, G. (1985). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge, MA: MIT Press.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, *22*, 133–187.

Fauconnier, G., & Turner, M. (2002). *The way we think*. New York: Basic Books.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*, 730–747.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*, 89–106.

Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1–55.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*, 1–26.

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gesture conveys substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, *2*, 67–74.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*, 1489–1522.

Gunter, T. C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters*, *372*, 52–56.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to pictue stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*, 1339–1450.

Hines, D. (1993). Effect of masked picture primes on semantic priming for easy- and difficult-to-name words. *Journal of General Psychology*, *120*, 149–169.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic processing. *Brain and Language*, *35*, 66–85.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47–61.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*, 259–276.

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology* (*vol. 3*, pp. 69–138). Greenwich, CT: AI Press Inc.

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory & Language*, *40*, 577–592.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*, 243–260.

Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–62). Cambridge, UK: Cambridge University Press.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, *7*, 54–59.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, *31*, 533–552.

Krauss, R. M., Morell-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, *61*, 743–754.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.

Kutas, M., & Hillyard, S. A. (1980). Event-related potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*, 99–116.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Kutas, M., & Hillyard, S. A. (1988). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*, 38–49.

McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*, 53–65.

Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M.,, Hinrichs, H., et al. (1999). IFCN standards for digital recording of clinical EEG. The international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement*, *52*, 11–14.

Oldfield, R. C. (1971). The assesment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Pratarelli, M. E. (1994). Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain and Cognition*, *24*, 137–157.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin*, *89*, 506–540.

Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research*, *5*, 54–62.

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*, 931–945.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, *40*, 160–164.

Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Review of Neurobiology*, *14*, 199–224.

van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.

Vanderwart, M. (1984). Priming by pictures in lexical decision. *Journal of Verbal Learning and Verbal Behavior*, *23*, 67–83.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*, 363–375.

Wu, Y. (2005). Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures. *CRL Newsletter*, *17*. Retrieved from http://www.crl.ucsd.edu/newsletter.

(Received December 7, 2004; Accepted August 4, 2005)