**Title**
Coordinated minds : how iconic co- speech gestures mediate communication

**Author**
Wu, Ying Choon Jane

**Publication Date**
2006

UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Coordinated Minds: How Iconic Co-speech Gestures Mediate Communication

A Dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Cognitive Science

by

Ying Choon Jane Wu

Committee in Charge:

Professor Seana Coulson, Chair
Professor Karen Emmorey
Professor Robert Kluender
Professor Marta Kutas
Professor Rafael Nuñez
Professor Martin Sereno

2006

The Dissertation of Ying Choon Jane Wu is approved, and it is acceptable in quality and form for publication on microfilm:

Chair

University of California, San Diego

2006

# TABLE OF CONTENTS

Signature Page …………………………………………………………….. ii

Table of Contents ……………………………………………………………. iv

List of Figures …………………………………………………………...... v

List of Tables………………………………………………………………… viii

Acknowledgements ………………………………………………………….. ix

Vita………………………………………………………………………...… xii

Abstract ……………………………………………………………………… xiv

Chapter 1: Introduction ……………………………………………………… 1

Chapter 2: Semantic Processing of Gestures……….…………………..……… 22

Chapter 3: Lexical Priming by Gestures……………………………………... 73

Chapter 4: Semantic Integration of Gestures and Speech………………………. 94

Chapter 5: Gesture Comprehension and Object Recognition…………………... 136

Chapter 6: General Discussion………………...………………………………... 195

# LIST OF FIGURES

## CHAPTER 2

Figure 2.1. Sample trial: A short cartoon segment was followed by a congruent or incongruent gesture video and then a probe word…………………………….. 33

Figure 2.2. 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……………………………………………………. 38

Figure 2.3. 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……………………………………………………………………. 41

Figure 2.4. 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………………………………………. 51

Figure 2.5. 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…………………………………………………………………………….. 53

Figure 2.6. Experiment 2: N400 elicited by related and unrelated words following congruous and incongruous contexts………………………………….. 56

Figure 2.7. Experiment 2: Isovoltage maps show the distribution of gesture congruency effects at 450 ms post-onset (Incongruent minus Congruent), and word relatedness effects at 400 ms post-onset (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………………………………………. 62

Figure 2.8. Examples of gestures used in Experiment 1 and 2. The speaker is portraying the dog's extended jaw, the opening box, and the dog's front legs…... 64

## CHAPTER 3

Figure 3.1. N400 elicited by related and unrelated words following spontaneously produced iconic co-speech gestures in Experiment 1. Data have been digitally filtered to remove frequencies greater than 15 Hz……………….. 80

Figure 3.2. N400 elicited by related and unrelated words following spontaneously produced iconic co-speech gestures in Experiment 2…………….. 84

Figure 3.3. Topography of N400 effect in Experiment 1 and 2………………….. 84

## CHAPTER 4

Figure 4.1. The speaker's hand configuration while uttering, "It's actually a double door."……………………………………………………………………….. 95

Figure 4.2. Experimental design. Cross-modal and speech-only picture probes occurred in related and unrelated trials………………………………………….. 102

Figure 4.3. Procedure. Discourse primes were followed by picture probes, and then target words………………………………………………………………. 110

Figure 4.4. Schematic diagram of scalp electrode sites………………………….. 111

Figure 4.5. ERPs recorded over central midline sites time-locked to the onset of picture probes……………………………………………………………………….. 115

Figure 4.6. Scalp topography of N300 relatedness effect (Unrelated minus Related) at 300 ms post-stimulus onset. The N300 effect was significant only in the cross-modal condition……………………………………………………… 116

Figure 4.7. The mean amplitude of N400 across all electrode sites (in microvolts) elicited by cross-modal and speech-only picture probes……………. 117

Figure 4.8. Spline maps depicting the distribution of the N400 effect (Unrelated minus Related) at 400 ms post-stimulus onset. For both cross-modal and speech-only conditions, values were normalized within a range of 1 and –1……. 119

**CHAPTER 5**

Figure 5.1. Trials included a short cartoon segment followed by either a dynamic or static gesture and then a probe word……………………………………………... 145

Figure 5.2. ERPs time-locked to the onset of static and dynamic gestures. Note that negative voltage is plotted up……………………………………………….... 155

Figure 5.3. ERPs elicited by related, unrelated, and unidentifiable photographs of common objects……………………………………………………………….. 157

Figure 5.4. A freeze frame extracted from a cartoon segment used in the experiment, and the subsequent static gesture……………………………………. 170

Figure 5.5. Normalized isovoltage maps (left) and current source density maps (right) computed from congruency effects elicited by dynamic and static gestures at the peak value of the congruency 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 congruency effects do not start until 400 ms post-stimulus.)…………………….. 172

Figure 5.6. Normalized isovoltage maps of N300 and N400 effects elicited by static gestures and object photos. Because the object photo N400 peaks earlier than the static gesture N450, these two effects are plotted at different time points……………………………………………………………………………... 176

Figure 5.7. Current source density constructed from isovoltage maps of the N300 and N400 congruency effects elicited by static gestures (left) and the relatedness effects elicited by photographs of common objects…………………. 177

# LIST OF TABLES

## CHAPTER 2

Table 2.1. Types of Iconic Gestures Used as Stimuli…………………………….. 31

Table 2.2. Experiment 1: Mean amplitude and standard error (in microvolts) of ERPs recorded over Fz, Cz, Pz, and Oz……………………………………….. 40

Table 2.3. Experiment 2: Mean amplitude and standard error (in microvolts) of ERPs recorded over Fz, Cz, Pz, and Oz……………………………………….. 52

Table 2.4. Mean Accuracy and Response Latencies in Classifying Probe Words.. 55

Table 2.5. Mean Accuracy in Classifying Probe Words………………………… 55

## CHAPTER 3

Table 3.1. Experiment 1: Analyses of Mean Amplitude of ERPs Elicited by Probe Words………………………………………………………………… 81

Table 3.2. Experiment 2: Analyses of Mean Amplitude of ERPs Elicited by Probe Words………………………………………………………………… 85

## CHAPTER 5

Table 5.1. Percentage of accurately classified words following congruent and incongruent gestures……………………………………………………………… 154

Table 5.2. Contrasts of mean amplitude of ERPs elicited by related, unrelated and NOID object photos between 300 and 400 ms post-stimulus………………. 159

Table 5.3. Contrasts of mean amplitude of ERPs elicited by related, unrelated and NOID object photos between 400 and 600 ms post stimulus………………. 162

Table 5.4. Contrasts of mean amplitude of ERPs elicited by related, unrelated and NOID object photos between 600 and 900 ms post-stimulus………………... 165

# ACKNOWLEDGEMENT

It has been a tremendous pleasure to work under the guidance of my advisor, Seana Coulson, as we navigated the many tide changes of ERP research. I am particularly grateful for her enthusiasm for the broad range of disciplines which constitute Cognitive Science and readiness to explore new horizons.

I would also like to thank the members of my committee – Karen Emmorey, Robert Kluender, Marta Kutas, Marty, Rafael Nuñez, and Martin Sereno. During my training in electrophysiological and neuroimaging methods, Marta Kutas and Martin Sereno have provided invaluable mentorship and resources.

A number of colleagues, including Morana Alac, David Groppe, Ruey-Song Huang, and Ayse Saygin, have shared their expertise throughout my graduate student career. I have also enjoyed the friendship of members of the Brain and Cognition Lab, including Jenny Collins, Chris Lovett, and Margeurite McQuire.

My decision to study Cognitive Science was inspired largely by my previous mentor, John Schuman, at UCLA. My decision to study language was influenced by John Haviland, who was at Reed College at the time. My decision to pursue graduate education was influenced by Dr. Saleh Abusaidi.

Finally, I would like to thank my mother, Christine Press, for her love and encouragement. I would also like to thank my father, David Wu, for his financial support during my undergraduate education. I would also like to thank my fiancé, Martin Klein, for his companionship in the final legs of this journey.

Chapter 2, in full, has been published in *Psychophysiology* 2005, Wu, Ying Choon; Coulson, Seana, Blackwell Press, 2005. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is currently in press for publication of the material as it appears in *Psychonomic Bulletin and Review*, Wu, Ying Choon; Coulson, Seana, Psychonomic Society Publications, 2006. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, has been accepted for publication of the material as it appears in *Brain and Language*, 2006, Wu, Ying Choon; Coulson, Seana, Elsevier, 2006. The dissertation author was the primary investigator and author of this paper.

CURRICULUM VITAE

## Education

| | |
|---|---|
| Sept 2000 –<br>2006 | University of California, San Diego<br>**Ph.D.** in Cognitive Science<br><br>Thesis committee chair: Dr. Seana Coulson |
| Sept 1998 –<br>June 2000 | University of California, Los Angeles<br>**M.A.** Applied Linguistics<br>Thesis title "The neurobiology of second language acquisition" |
| Aug 1988 –<br>June 1993 | Reed College<br>**B.A.** English Literature |

## Journal Articles

Wu, Y. C. & Coulson, S. (accepted). How iconic gestures enhance communication: an ERP study, *Brain and Language*.

Wu, Y. C. & Coulson, S. (in press). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin and Review*.

Wu, Y. C. & Coulson, S. (2005). Meaningful gestures: Electrophysiological evidence of gesture comprehension. *Psychophysiology*, 42 (6), 654-667.

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

Coulson, S. & Wu, Y. C. (2005). Hemispheric asymmetries in joke comprehension. *Journal of Cognitive Neuroscience*, 17 (3), 494-506.

## Abstracts

Wu, Y.C. & Coulson, S. 2004. How Iconic Gestures Enhance Communication: An ERP Study. A supplement of the Journal of Cognitive Neuroscience, p. 171.

Wu, Y.C. & Coulson, S. 2004. Is that a Meaningful Gesture: Electrophysiological Indices of Gesture Comprehension. A supplement of the Journal of Cognitive Neuroscience, p. 31.

Wu, Y.C. & Coulson, S. 2003. Hitting a Home Run: Joke Processing in the Right Hemisphere. A supplement of the Journal of Cognitive Neuroscience, p. 176.

## Invited Presentations

Wu, Y.C. (2006, July). Gestures worth a thousand words. Neurocognition of Gesture Workshop conducted at the Max Planck Institute, Leipzig, Germany.

Wu, Y. C. (2003, July). Frame-shifts in action: What spontaneous humor reveals about joke comprehension. Symposium conducted at the 8th International Cognitive Linguistics Conference, University of La Rioja, Spain.

Wu, Y. C. (2002, April). Embodied action, embodied knowledge: What gesture reveals about conceptual processes. Talk presented at the Empirical Methods in Cognitive Linguistics Workshop, Cornell University, New York.

## Honors and Awards

*Fellowships*

| | |
|---|---|
| 2005-2006 | Predoctoral Fellowship, Institute for Neural Computation/ National Institute of Health |
| 2004-2005 | Predoctoral Fellowship, Center for Research in Language/ National Institute of Health |
| 1999-2002 | NSF Graduate Fellowship Award |
| 1998-1999 | Graduate Opportunity Fellowship, UCLA |
| 1993-1994 | Oregon Literary Fellowship, Oregon Institute for the Literary Arts |

*Honors*

| | |
|---|---|
| 2002, 2003 | Superior Teaching, Department of Cognitive Science, UCSD |
| 2001 | Excellence in Teaching, Department of Cognitive Science, UCSD |

## Professional Background and Teaching

Summer 2005
Instructor, John Hopkins University
Cognitive Psychology

Spring 2004
Teaching Assistant, Department of Cognitive Science, UCSD

CogSci 101C: Language and Meaning

Winter 2003, 2004
Teaching Assistant, Department of Cognitive Science, UCSD
CogSci 101B: Learning, Memory, and Attention

Spring 2001
Teaching Assistant, Department of Cognitive Science, UCSD
CogSci 101C: Language and Reasoning

1995-1997
Medical Interpreter (Russian), Multnomah County Health Services; Portland, OR

1994-1995
Bilingual instructor, International Learning Program; Portland, OR

1993-1994
English Language Instructor, Ivanovo School District / Ivanovo State University, Ivanovo, Russia

## Professional Activities, Memberships, and Services

*Membership*
Cognitive Neuroscience Society, 2003-present

*Organizational positions and service*
Organizer for Cognitive Science WA Series: 2004, 2005
Organizer for Cognitive Science Distinguished Speaker Series: 2005-2006

*Mentoring*
Spring through Fall 2006: Anjana Khosla
Fall through Spring 2004 and 2005: Philip Lai, Kelly McCormick
Winter 2004: Sarah Smolley
Fall 2002 through Spring 2003: Shabaik, Salma
Winter and Spring 2002: Chiv, Mouy; Fiske, Laurel; and Nam, Cindy
Spring 2001: Yessis, Marissa

## ABSTRACT OF THE DISSERTATION

# Coordinated Minds: How Iconic Co-speech Gestures Mediate Communication

by

Ying Choon Jane Wu

Doctor of Philosophy in Cognitive Science

University of California, San Diego, 2006

Professor Seana Coulson, Chair

Iconic co-speech gestures are spontaneous body movements produced in coordination with speaking. Concurrently with the word *platter*, for example, a speaker might trace an oval in the air, showing something about the shape of the object that he is conceptualizing. Chapter 2 investigates fundamental cognitive processes mediating gesture comprehension by comparing ERPs elicited by contextually congruent and incongruent gestures.

Because gestures are not part of a conventionalized symbolic system, researchers have argued that they do not convey substantive content on their own; rather, their meaning is driven by speech that accompanies them. Chapter 3 tests these claims by measuring real time semantic activations prompted by iconic gestures presented in the absence of supporting context.

To assess how gestures affect comprehension of discourse, EEG was recorded as healthy adults viewed short segments of spontaneous discourse involving both gestures and verbal utterances (Chapter 4). Discourse segments were followed either by related picture probes, which corresponded with information made available both in speech and gesture (Cross-modal Matches), or in speech alone (Speech-only Matches), or by unrelated controls. By comparing brain response to Cross-modal and Speech-only Matches, it is possible to assess the specificity of semantic activations during the integration of speech and gestures.

Finally, Chapter 5 investigates whether iconic gestures engage object recognition processes implicated in the comprehension of conventionally depictive representations, such as photographs. "Static gestures" were created by extracting from each dynamic gesture a single freeze frame which made visible critical information about the speaker's intended meaning. EEG was recorded as participants viewed static and dynamic gestures, as well as photographs of common objects. The distribution and time course of ERP effects elicited by these stimulus types were compared.

# CHAPTER 1: GENERAL INTRODUCTION

In conversation, speakers frequently move their hands and bodies in a way that corresponds with their meaning. In describing a past encounter with a rattlesnake, for example, a man kneels and points towards the grass, rotating his extended arm and finger in a circular motion, while uttering, “The snake was right there.” Purely on the basis of this gesture, a listener might infer that the speaker is describing a fairly small, elliptically bound spatial extent near his feet. On the basis of both the utterance and the gesture, the inference might become available that the snake was coiled when the speaker happened to step down beside it. Significantly, the notion of a coiled shape is overtly specified neither in the speaker’s words, nor in his gesture. Rather, it is an emergent property of the integration of stored knowledge about snakes, activated by the utterance, with visuo-spatial percepts prompted by the gesture.

The idea that during conversation, listeners integrate analogue information in gestures with propositional information in speech is both intriguing and controversial. It is intriguing because in themselves, gestures are merely extemporaneous, fleeting movements of the hands and body. That people regularly communicate fairly complex concepts through this medium – such as a coiled snake – is a property of human behavior that is beginning to attract attention from cognitive neuroscience.

On the other hand, the idea of speech and gesture integration is controversial due to the somewhat indeterminate status of most co-speech gestures as a communicative medium. While it is generally agreed that these gestures differ from explicit actions (such as coughing or jumping) in that they are fundamentally representational in nature, the extent of their representational capacity is still poorly understood. Unlike sign languages used within deaf communities, most co-speech gestures are systematic neither in form nor meaning – that is, standards of well-formedness governing language production do not appear to characterize gesture production, and a single gesture can be used to denote an almost endless range of concepts. In fact, behavioral research has yielded findings both in favor of and contrary to the view that co-speech gestures convey substantive semantic content to listeners at all.

The present body of research will address the question of how gestures affect communication by studying the kinds of information that they activate during real-time comprehension. Additionally, we will investigate cognitive and neural systems mediating the comprehension of gestures. Because gestures and speech share close rhythmic, semantic, and developmental links, it is possible that both of these channels engage a common underlying substrate (Bates & Dick, 2002). Alternatively, because gestures are apprehended mainly through the visual modality, and because they often are realized through movements of the hands and upper body, understanding them may involve processes recruited during object recognition or process recruited in the comprehension of meaningful actions and biological movement.

The first question to be addressed is whether understanding gestures engages cognitive processes similar to those engaged by more conventionally meaningful representations such as pictures and words. Secondly, we will investigate information activated by gestures. Chapter 3 will study semantic activations prompted when gestures are presented independently of accompanying speech or other forms of contextual support. Chapter 4 will study the specificity of semantic activations prompted by gestures which add information above and beyond what is made available in accompanying speech. Finally, Chapter 5 will compare brain systems mediating the comprehension of gestures and common objects.

### *Semantic processing of gestures*

The many types of behavior classifiable as gesture exhibit heterogenous semiotic properties – that is, different types of relationships obtain between the gestural form, its meaning, and its real-world extension. McNeill (1992; 2005) illustrates how several distinct classes of body movement can be distinguished in terms of the degree to which they are associated with conventionalized meanings, and the degree to which speech obligatorily accompanies them. On one end of the spectrum, gesticulation comprises those gestures which occur exclusively during the course of speaking, and whose meaning is determined largely by the context in which they are produced. This category includes iconic gestures, which derive their capacity for signification from relational similarities between properties of the gesture (e.g., hand shape, location, trajectory, rate of motion, and so forth) and a highly schematic conceptualization of properties of the referent. It also encompasses metaphoric gestures, which use visuo-spatial cues to represent abstract concepts, such as time or thought. Additionally, beats, which are small hand flicks coordinated with speech prosody, also belong to this category.

At the opposite end of the spectrum are full-fledged sign languages such as American Sign Language (ASL). In contrast to gesticulation, hand signs exhibit linguistic structure and activate a more stable system of meanings (Emmorey, 1999).

Further, they are typically produced in the complete absence of speech. In between the extremes of sign language and gesticulation are emblems (e.g., hitchhiking sign or thumbs up) and deictic pointing. Like the lexical items in sign language, emblems express fully conventionalized meanings; however, they are frozen forms which do not encode linguistic information. They are produced either in the presence or absence of accompanying speech. Deictic gestures also occur both with and without speech. They are closer to gesticulation than emblems, though, since context plays a critical role in determining their meaning.

From this brief taxonomy, it is apparent that co-speech gestures encompass a wide range of communicative behaviors. Further, converging lines of experimental evidence suggest at least some degree of separation in the neural resources engaged by each type. For instance, in the case of sign language, dissociations have been reported in aphasic signers sustaining left hemisphere brain damage whereby language function is impaired in the face of preserved capacity for non-linguistic pantomimic and iconic gesture (Corina et al., 1992; Marshall, Atkinson, Smulovitch, Thacker, & Woll, 2004). In one case, an aphasic deaf man was shown to experience difficulty naming objects in British Sign Language (BSL), but not gesturing their use or appearance (Marshall et al., 2004). Significantly, this dissociation held true even for objects whose BSL name resembled typical gestures produced by deaf and hearing controls in response to those objects. In sum, the capacity to communicate through gesture and sign language appear to rely upon distinct neural substrates.

Functional magnetic resonance imaging has also revealed differences in neural activations elicited by sign language and gesture. In one study, fMRI was used to measure BOLD signal changes during the presentation of BSL sentences as compared with strings of emblems used at racecourses to communicate odds on horses (Tic Tac) (MacSweeney et al., 2004). Relative to baseline, where the signer appearing in the other conditions was shown at rest, BSL and Tic Tac both engaged fairly similar cortical networks in hearing and deaf signers, who were instructed to monitor for infrequently occurring semantic anomalies. Subtracting signal elicited by Tic Tac from that elicited by BSL, however, revealed subtle differences in the neural signature prompted by each stimulus type, particularly in the left hemisphere. Activations attributable exclusively to BSL were observed in left posterior perisylvian regions including the superior temporal gyrus and sulcus for both deaf and hearing signers, as well as the supramarginal gyrus for deaf signers only. These regions were proposed to be sensitive to phonetic and phonological properties of BSL that were not present in Tic Tac.

The studies reviewed thus far suggest distinct neural substrates mediating sign language use and iconic or pantomimic gesturing, given dissociations in the effect of left hemisphere brain damage on these two behaviors. On the other hand, understanding emblems, such as those comprising Tic Tac, appears to engage left hemisphere cortical networks which overlap, but are not co-extensive, with those responsible for sign language. One caveat with regard to the MacSweeney et al (2004) study should be noted, however – namely, that none of the participants were familiar with Tic Tac gestures. Thus, their findings may not necessarily be reflective of brain activity prompted during the observation of commonly recognizable emblems.

Research specifically involving well-known emblems (e.g., OK sign) has also suggested both commonalities and differences in the cognitive systems mediating spoken language and emblem comprehension. Gunter and Bach (2004) compared event-related potentials (ERPs) elicited by common, consistently interpreted emblems and meaningless hand configurations. Relative to common emblems, meaningless items resulted in more negative ERPs over anterior electrode sites between 300 and 400 ms post-stimulus (N300) and over posterior electrode sites between 450 and 550 ms post-stimulus (N400). This pattern of results is concordant with studies reporting larger N300 and N400 in response to pictures of unidentifiable objects relative to nameable ones (W. B. McPherson & P. J. Holcomb, 1999), as well as with studies reporting larger N400 in response to pseudowords relative to legal ones (Bentin, 1987).

Only reported in response to pictorial stimuli, the N300 is thought to reflect image-specific aspects of object recognition, with incongruous or non-canonical images eliciting N300 with greater amplitude than those which are easier to identify. By contrast, the N400, which was originally discovered in response to semantically incongruous sentence final words (Kutas & Hillyard, 1980b) has been observed in response to a wide range of contentful stimuli in addition to words, including pictures (Barrett & Rugg, 1990), ASL hand signs (Neville et al., 1997), action videos (Sitnikova, Kuperberg, & Holcomb, 2003), and environmental sounds (Plante, Van Petten, & Senkfor, 2000). Because the amplitude of the N400 is inversely correlated with the degree to which a stimulus fits its context, this component is thought to index the integration of semantic activations induced by a given stimulus with other contextually activated information.

The finding that emblems result in smaller N400 than nonsensical hand configurations suggests that their meaning is processed in a manner similar to meanings made available by words. On the other hand, the N300 effects observed in this study also indicate that emblems – at least when presented in pictorial form – recruit image-specific semantic processes as well. These outcomes are in keeping with the consistent form-meaning mappings which are a hallmark of this class of gesture.

Until recently, it was largely unknown whether less conventionally meaningful classes of gesture, such as iconic gestures, engage semantic processes analogous to those recruited by emblems. Unlike emblems, iconic gestures do not constitute a stable system of meanings. The same circular pointing gesture accompanying the snake story, for instance, could be used to describe a wide range of phenomena in different contexts, including the location of a plant, or an area to be marked with spray paint. Because the significance of any given iconic gesture depends crucially on the context in which it is produced, it is possible that this type of semiotic resource recruits semantic integration processes that differ from those invoked by emblems, words, and other systematically meaningful stimuli. Such a possibility is consistent with behavioral research which reports minimal effects of iconic gestures on discourse comprehension (Krauss, Dushay, Chen, & Rauscher, 1995; Rime, 1982).

In response to this issue, Chapter 2 will describe a set of experiments which assess the real time comprehension of iconic gestures. By measuring ERPs elicited by spontaneously produced iconic gestures presented without their original accompanying speech, it was possible to test whether these items on their own elicit neural signals associated with semantic integration. To redress the loss of contextual support due to the removal of concomitant speech, each gesture stimulus was preceded by either a congruent or incongruent segment of animated cartoon. In congruous cases, the preceding cartoon segment had been the topic of the speaker's speech and gestures when the raw material for the stimuli was video recorded. In incongruous cases, a different cartoon segment was used. If iconic gestures engage semantic integration processes, then this manipulation of gesture congruency is predicted to result in an N400-like context effect.

### *Semantic activations elicited by iconic gestures*

Some researchers have argued that iconic gestures are inherently vague, and their meaning is determined largely by the speech which accompanies them. One author writes, "…it may be that much of the gesture's meaning is illusory. In the absence of speech, the very same gesture's meaning can be quite opaque, communicating little, if anything"; (Krauss, Morell-Samuels, & Colasante, 1991). This view is supported by a number of studies in which participants attempt to guess the meaning of iconic gestures when speech context is absent or reduced. For instance, Hadar and Pinchas-Zamir (2004) presented video taped segments of spontaneous discourse in three different formats – either silently without audio sound track; silently with the accompanying speech shown simultaneously in text format; or with both text and audio sound track. In both the text and the sound track, the lexical affiliate corresponding with the speaker's gesture was removed, and participants were instructed to choose from a set of five possible words the one which best described the gestures. The response set included the actual lexical affiliate (e.g., snake), a distractor word semantically related to the lexical affiliate (e.g., lizard), a distractor word related to visual properties of the referent denoted by the lexical affiliate (e.g., stick), as well as two unrelated choices.

Even in the most informative text+audio condition, participants selected the lexical affiliate of iconic gestures only with 40% accuracy. Though well above chance (20%), this accuracy rate led the authors to conclude that participants' interpretations of iconic gestures were "not very specific (p. 210)." This view is in keeping with a similar study (Feyereisen, van de Wiele, & Dubois, 1988) demonstrating that without verbal contextual support, viewers were more likely to match iconic gestures with plausible or unrelated distractors than their correct lexical affiliate. Likewise, Krauss et al (1991) demonstrate that without accompanying speech, gestures were interpreted in a manner judged consistent with their lexical affiliate at a rate of only 12% above chance. On the remaining trials, interpretations of gestures were judged more consistent with a different word that was not part of the original accompanying speech. The authors conclude, "…though gestures may be a guide to what is being conveyed verbally, it would be difficult to claim on the basis of these data that they are a particularly discriminating guide"; (Krauss et al., 1991)."

In contrast, a slightly different set of assessment techniques has led other researchers to the opposite conclusion. Beattie and Shovelton (1999b; 2002) video-taped naïve individuals describing comic strips, and selected instances of iconic gesture from these narratives to be presented without audio sound track to university undergraduates. Participants responded to comprehension questions regarding various features of objects being described, including their size, location, shape, movement, speed, and identity. For some categories, such as movement, size, and location, the accuracy rate of participants' responses was well above chance – particularly for gestures depicting events from an agentive perspective. This finding suggests that iconic gestures are capable of conveying substantive semantic content independently of speech, and could therefore conceivably add information above and beyond what is made available in speech.

One advantage of the methods employed by Beattie and Shovelton is that open-ended comprehension questions provide a more sensitive measure of information communicated by gestures than requiring participants to select a lexical affiliate. However, all of these studies are problematic in that they interpret their data to confirm pre-existing biases without considering alternatives. For instance, on the one hand, Krauss et al argue that iconic gestures are semantically austere given the finding that viewers' construals of such gestures were consistent with the meaning of their lexical affiliates at a rate only 12% better than chance. On the other hand, Beattie and Shovelton propose that iconic gestures are communicatively rich, given the fact that after viewing them in the absence of speech, participants responded to comprehension questions with mean accuracy rates ranging from 11% to 36%. Because Krauss et al used a forced choice classification paradigm, whereas Beattie and Shovelton used a questionnaire-based approach, it is difficult to compare their outcomes. However, the juxtaposition of their conclusions illustrates that their methods do not provide a conclusive index of gesture meaningfulness – as small improvements in comprehension relative to baseline can be construed either as evidence in favor of or in opposition to the capacity of gestures for signification.

Rather than attempting to quantify the communicative value of iconic gestures, the experiments described in Chapter 3 investigate whether they prime related words. Given the studies described above, iconic gestures do not appear to map consistently to the lexical items with which they co-occur. Rather, we hypothesize that they weakly activate semantic information related to a potentially broad range of referents whose visuo-spatial properties are consistent with the gestures' features, and that these semantic activations will be evident in ERPs elicited by visually presented words either related or unrelated to gesture primes. As a first step to testing this hypothesis, we presented viewers with the same silent gesture stimuli described in Chapter 2, but removed preceding cartoon contexts. Each gesture was paired with a related or unrelated probe word. If iconic gestures activate related concepts, evidence of lexical priming is expected.

### ***Semantic integration of speech and gesture***

Over the past decade, gestures have been the focus of considerable attention from researchers investigating their role in thought and communication. McNeill (1992) has proposed that gestures constitute an integral component of language, providing an analogue medium whereby speakers may convey information which does not readily lend itself to linguistic encoding. Evidence in support of a functional link between speech and gesture can be found in their temporal and semantic alignment.

Speakers produce 90% of gestures while actively engaged in producing speech (McNeill, 1992). When speech is disrupted, as in the case of stuttering, gesturing is also attenuated (Mayberry & Jacques, 2000). Further, gestures tend to express meanings which correspond with the speech that they accompany.

Just as speech and gesture are thought to comprise a cooperative system in speakers, concurrent information made available through these modalities has been proposed to be integrated by listeners. Cassell, McNeill, & McCullough (1999) presented participants with videotaped segments of a confederate describing an animated cartoon. In some cases, the confederate produced gestures which agreed with the semantic content of their speech (e.g., saying, "And then she offers him a penny," while making an offering gesture in the direction of the listener). In other cases, the same utterance was deliberately accompanied by a conflicting gesture (e.g., the speaker makes an offering gesture to himself). After each segment of narrative, participants were instructed to describe what they had understood thus far.

Sensitivity to information expressed in mismatching gesture stimuli was evident both in the speech and gestures produced over the course of retelling. Listeners produced many more errors in their accounts of the narratives after viewing mismatching gestures as compared to matching ones. For instance, one participant describes the video segment containing the mismatching offering gesture as follows:

> Granny sees him and says, "Oh what a nice little organ grinder," [and she] gets – [goes to give hi]m [a penny] – a little monkey excuse me (p. 22).

Over the course of the utterance, the participant produced three gestures (corresponding with the bracketed portions of her speech) – first, making an offering gesture in the direction of the listener, then towards herself, and finally, towards the listener once more.

Of particular importance in this example is the fact that conflicting information conveyed by the original mismatching stimulus gesture was expressed jointly in the speaker's own utterance and gestures. In other words, input transmitted through the gestural channel was represented in a format accessible for subsequent expression both by verbal and gestural means. This finding has been interpreted as evidence in favor of the view that in comprehension, information from speech and gestures is integrated in long-term memory, rather than stored as distinct, independent representations.

Other studies also support this view. Goldin-Meadow, Wein and Chang (1992) presented adults with videotaped vignettes in which children responded to Piagetian conservation tasks. As in the previously described experiment, half of the vignettes showed children whose speech and gestures matched (e.g., saying that a vessel was tall while indicating its height in gesture). The remaining stimuli showed children who spontaneously produced gestures that did not match their speech (e.g., describing the height of a vessel in speech, but indicating its width in gesture). After each stimulus vignette, adult subjects were asked to explain the reasoning underlying the child's response, and their own speech and gestures were recorded and analyzed.

Relative to matching trials, adult participants produced more of their own speech-gesture mismatches when explaining vignettes involving mismatching gestures. Further, participants' explanations of mismatching vignettes included information expressed only in the children's gestures. Subsequent research has corroborated these findings both in children (Kelly & Church, 1998) and adults (Alibali, Flevares, & Goldin-Meadow, 1997). It has also been shown that viewers can successfully glean information from gestures even while observing children performing conservation tasks "live" rather than on video (Goldin-Meadow & Sandhofer, 1999).

When viewed in concert, these studies provide considerable evidence in support of the communicative value of co-speech gestures. Additional research has demonstrated that gestures can enhance comprehension of discourse. However, see Krauss et al (1995) for a different view. At the level of dyadic exchanges, listeners have been shown to produce more target-like drawings of abstract designs in response to verbal descriptions when their interlocuters' spontaneous co-speech gestures are also visible (Graham & Argyle, 1975). Similarly, when watching video recordings of spontaneous descriptions of comics or films, listeners exhibit more accurate comprehension of the event under description – especially with regard to its visuo-spatial properties – when the speaker's gestures were visible as compared to when only the audio soundtrack was played (Beattie & Shovelton, 1999a, 1999b; Rogers, 1978).

In other types of communicative interchanges, such as classroom instruction, gestures also benefit understanding. Valenzeno, Alibali, and Klatzky (2003) found more accurate post-test performance from pre-school children presented with a video-taped lesson on symmetry when tracing and pointing gestures were used as compared to a version of the lesson in which the teacher simply kept her arms at her side.

Likewise, Church, Ayman-Nolley, and Estrada (2004) reported that elementary school children who were either proficient in English, or learning English as a second language, both exhibited improved performance in response to a videotaped lesson in English on Piagetian conservation tasks when the explanations were accompanied by a rich array of representational gestures as compared to when they were accompanied only by minimal pointing. Finally, Singer and Goldin-Meadow (2005) showed that elementary school children benefited from lessons on math equivalence problems when the instruction involved gestures that conveyed additional information beyond what was expressed in speech. Post-test scores were higher in response to this type of instruction than lessons which involved speech and matching gestures or speech produced without accompanying gestures.

To summarize, the past three decades of behavioral research demonstrate that co-speech gestures contribute substantively to listener comprehension. However, little is known about the cognitive and neural systems mediating this phenomenon. We hypothesize that co-speech gestures activate image-specific information which enables listeners to formulate fairly specific conceptual representations of the speaker's intended meaning. To address this issue, Chapter 4 will describe a set of experiments in which we recorded EEG from healthy adults as they viewed short segments of spontaneous discourse involving both gestures and verbal utterances. Discourse segments were followed either by related picture probes, which corresponded to information made available both in speech and gesture (Cross-modal Matches), or in speech alone (Speech-only Matches), or by unrelated controls (Mismatches). If language users integrate propositional information in the speech stream with analogue information in gestures, we predict images consistent with information conveyed cross-modally to be apprehended more readily than those consistent with speech alone.

### *Cognitive processes mediating gesture and picture comprehension*

As suggested by Chapter 4, iconic gestures may enable listeners to represent visuospatial properties of the speaker's intended meaning more precisely. Chapter 5 will explore this idea further by investigating commonalities and differences between the semantic processing of pictures and gestures. A paradigm will be used similar to that employed by Ganis, Kutas, and Sereno (1996), who measured ERPs elicited by congruent and incongruent sentence completions which were presented either in word or pictorial form. The researchers found that while both incongruent words and pictures elicited larger N400 than congruent ones, this effect was more prominent over anterior electrode sites for pictures, while words resulted in greater N400 effects over posterior electrodes. The similarities in size and time course of these N400 effects were interpreted as evidence for functional overlap in the semantic processing of words and pictures. On the other hand, the different distributions of the N400 effects were construed as evidence that understanding words and pictures engages at least partially non-overlapping neural systems.

Chapter 5 will describe a set of within subject experiments comparing ERP effects resulting from congruency manipulations of gestures and photographs of common objects. Using stationary "snapshots" of gestures affords the opportunity to assess resulting brain activity by means of stimuli which share the static quality of still photos and line drawings used in extant research on the picture N400. We created "static gestures" by extracting from each dynamic gesture movie clip a single freeze frame that made visible critical information about the speaker's intended meaning. Freeze frames were preceded by the same congruous and incongruous cartoon contexts described in Chapter 2. Photographs were presented as congruous or incongruous prime-target pairs constructed from the corpus of stimuli used in McPherson and Holcomb (1999).

In keeping with McPherson and Holcomb (1999), we expected incongruous pictures to result in larger N300 and N400 relative to congruous ones. Given results reported in Chapter 2, we also expected static gestures to elicit N400-like congruency effects. If static gestures engage image-specific processes similar to those triggered by photographs of emblems (Gunter & Bach, 2004) and other kinds of static image-based representations, N300 congruency effects are expected as well. Alternatively, it is possible that the iconicity of gestures is too schematic to affect object recognition processes indexed the N300. Finally, by comparing the distribution and current source density of ERP effects elicited by static gestures and photographs of common objects presented in this experiment, it will be possible to draw inferences about the degree of commonality in the underlying cortical systems recruited by gestures and more directly recognizable images.

## Works Cited

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.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS. *Trends in Cognitive Sciences, 4*(7), 267-278.

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

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *developmental Psychobiology, 40*(3), 293-310.

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*(2), 179-192.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language, 31*(2), 308-327.

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), 1-33.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism, 7*(4), 303-319.

Corina, D. P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., & O'Grady-Batch, L. (1992). Dissociation between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Language, 43*(3), 414-447.

Emmorey, K. (1999). Do signers gesture? In R. Campbell (Ed.), *Gesture, speech, and sign*. Oxford: Oxford University Press.

Feyereisen, P., van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech. *European Bulletin of Cognitive Psychology, 8*(3-25).

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*(2), 89-106.

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

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge thorugh gestures: Using children's hands to read their minds. *Cognition and Instruction, 9*(3), 201-219.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology, 10*, 57-67.

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

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture: Implications for gesture classification and function. *Journal of Language and Social Psychology, 23*(2), 204-214.

Kelly, S. D., & Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*(1), 85-93.

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*(5), 743-754.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-305.

MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., et al. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage, 22*, 1605-1618.

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., & Woll, B. (2004). Aphasia in a user of British Sign Language: Dissociations between sign and gesture. *Cognitive Neuropsychology, 21*(5), 537-554.

Mayberry, R. I., & Jacques, J. (2000). Gesture production during stuttered speech. In D. McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge University Press.

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

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: 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.

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating american sign language: Effects of sensory experience and age of acquisition. *Brain and Language, 57*(3), 285-308.

Petrides, M., & Pandya, D. N. (1997). Comparative architectonic analysis of the human and the macaque frontal cortex. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (Vol. 9, pp. 17-58). New York: Elsevier.

Plante, E., Van Petten, C. K., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia, 38*, 1669-1684.

Rime, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology, 12*(2), 113-129.

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

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85-89.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teacher's gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology 29*(187-204).

# CHAPTER 2: SEMANTIC PROCESSING OF GESTURES

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 (Adam 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 et al., 1997; Cassell et al., 1999; Goldin-Meadow & Sandhofer, 1999; Kelly, Barr, Church, & Lynch, 1999). In one such experiment, gestures which 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). 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 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 et al., 1995). Other behavioral studies suggest that listeners rely heavily on semantic information conveyed through speech in order to attribute meaning to accompanying gestures (Krauss et al., 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; P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. 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, 1980a, 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; W. B. McPherson & P. J. Holcomb, 1999). Moreover, just as pseudo-words elicit larger N400s than unrelated words (Holcomb, 1988), unrecognizable images elicit larger N400s than recognizable ones (P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. 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 et al., 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 (W. B. McPherson & P. J. 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 pre-stroke 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 2.1). In some cases, gestures highlighted central relations depicted in the cartoon; in others, they emphasized fairly incidental details.

**Table 2.1.** 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 | path of an arrow in flight | length of a bridge |
| | applying glue | chomping jaws | shape of a panel |
| | adjusting a robot | impact of a collision | location of buttons |
| | lifting a lid | a dog striding | shape of a lid |

Total = 160

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 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 2.1 shows a schematic of a sample trial. Four additional trials were used in a practice block at the outset.

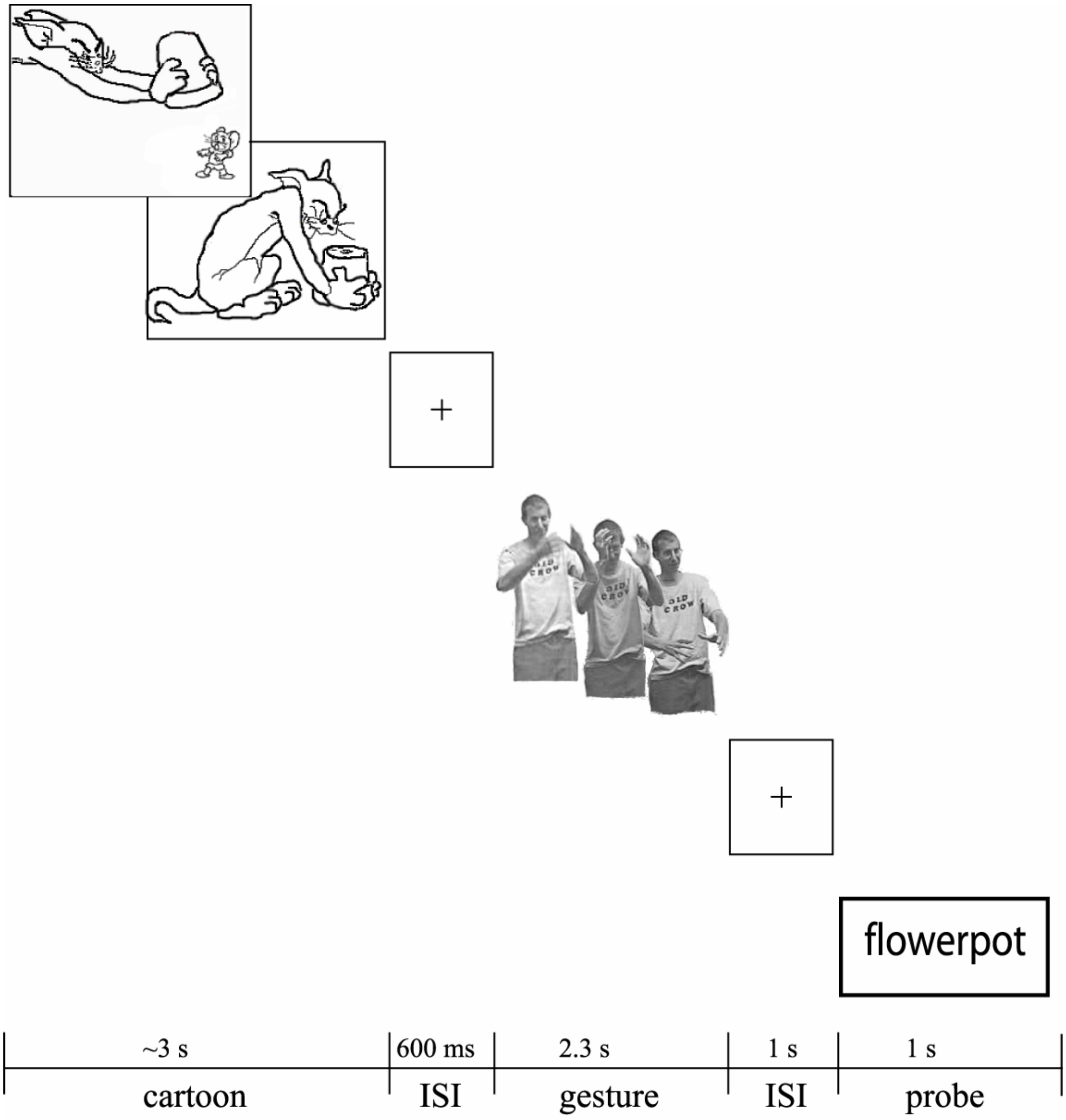


**Figure 2.1.** Sample trial: A short cartoon segment was followed by a congruent or incongruent gesture video and then a probe word.

### *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 re-referencing, below the right eye for monitoring blinks, and at the outer canthi for monitoring eye movements. All electrodes were referenced online 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-second window post-stimulus onset (5% congruous and 8% incongruous trials lost due to trimming), and a 4-second 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 offline 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 ±16 μV were flagged for contamination from blinking. Eye movements were monitored by means of a bi-polar 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 ±6 μV 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 post-stimulus, 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 pre-stimulus 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

---

[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

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 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).

---

from 100 to 1200 ms. Statistically reliable effects of gesture congruency were observed continuously from 300 to 1200 ms.

## 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 second 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 2.2). Overlap with the larger onset negativity may have

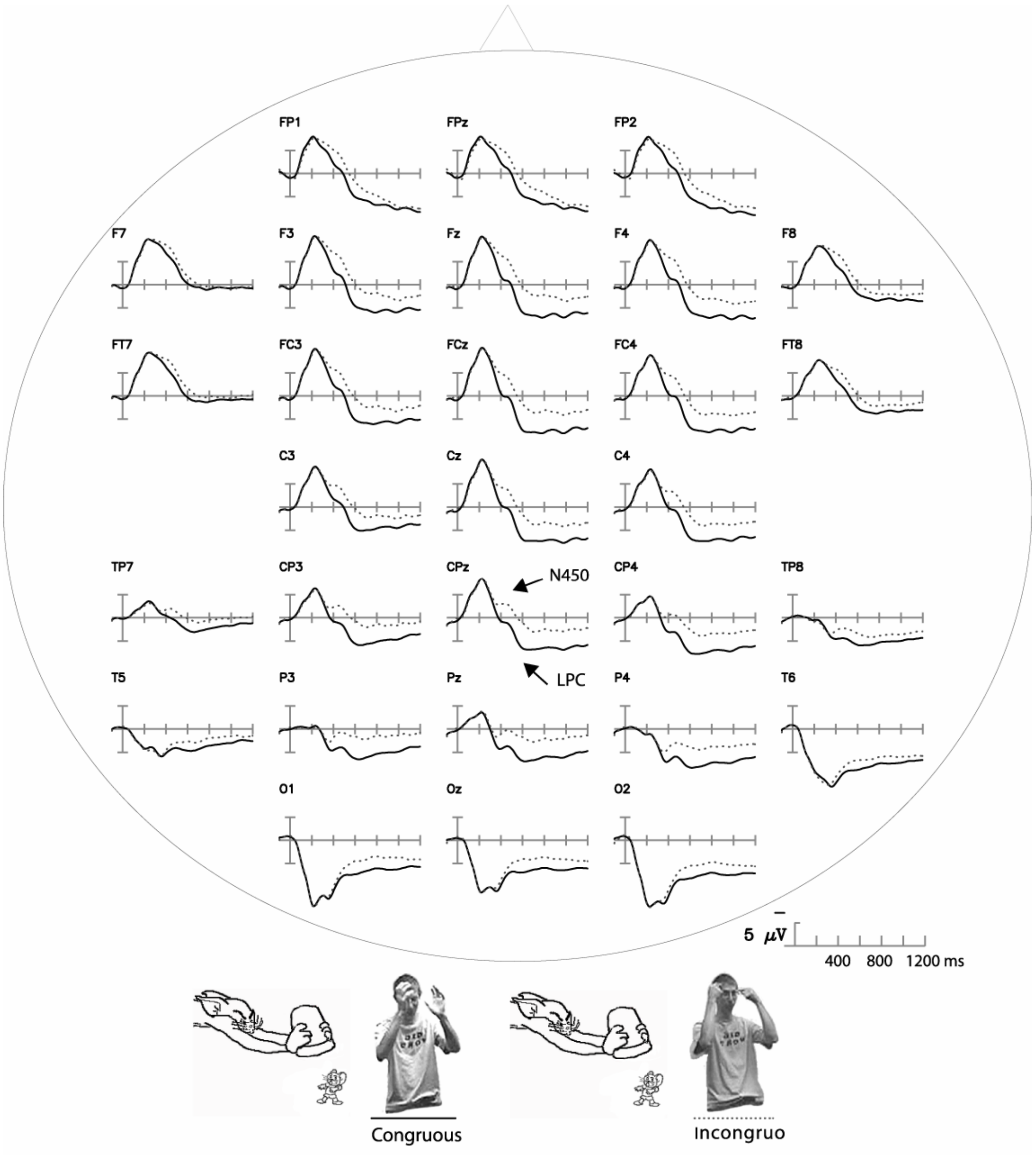


**Figure 2.2.** Experiment 1: ERP responses time-locked to the onset of congruous and incongruous gestures and extending for 1200 ms. Time zero corresponds with the onset of gesture clips.

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, ε = .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 2.2 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, ε = 0.125), and 900 to 1200 ms post-stimulus (Congruency main effect: F[1,11]= 27.0, p < .0005); Congruency × Electrodes Interaction F[28,308]= 4.6, p < .01, ε = 0.1), again with maximal effects at the fronto-central midline.

**Table 2.2.** 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 |

### *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< 1.5, n.s.

*ERPs.* As shown in Figure 2.3, 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).

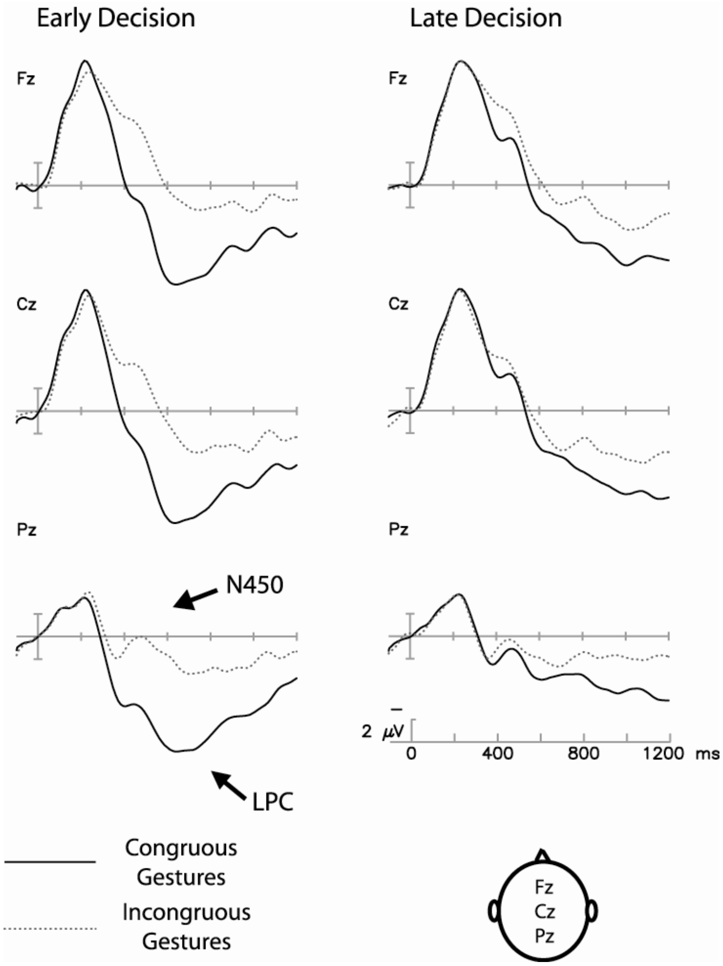


**Figure 2.3.** Experiment 1: Gesture ERPs sorted according to subjects'median decision times. The N450 effect occurs primarily in cases where the relationship between cartoon and gesture was readily apparent. Congruency effects occur later for late decision trials. Time zero corresponds with the onset of gesture clips.

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([,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 post-stimulus 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 post-stimulus, 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, ε = .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 post-stimulus (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 studies 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; P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. Holcomb, 1999; Sitnikova et al., 2003; West & Holcomb, 2002).

Although incongruous gestures elicited more negative ERPs even earlier, between 300 and 400 ms after stimulus, 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 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 post-stimulus, 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.

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).

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 four 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 by 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).

In order 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, s.d .=.0 57 (incongruent: 1.3, SD = 0.37); remaining trials were binned as Moderately Congruent items, with a mean rating of 3.8, SD0= 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.

---

[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.

## Results

### *Gestures*

Both congruous and incongruous items elicited a large negative onset potential, peaking around 230 ms (see Figure 2.4). 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 F's < 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 to 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 to 1200 ms window (main effect: $F[1,11] = 22.0$, $p < .005$; Congruency x Electrode Site interaction: $F[28,308]= 3.3$, $p < .05$, $\varepsilon= .15$). As shown in Table 2.3, 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.

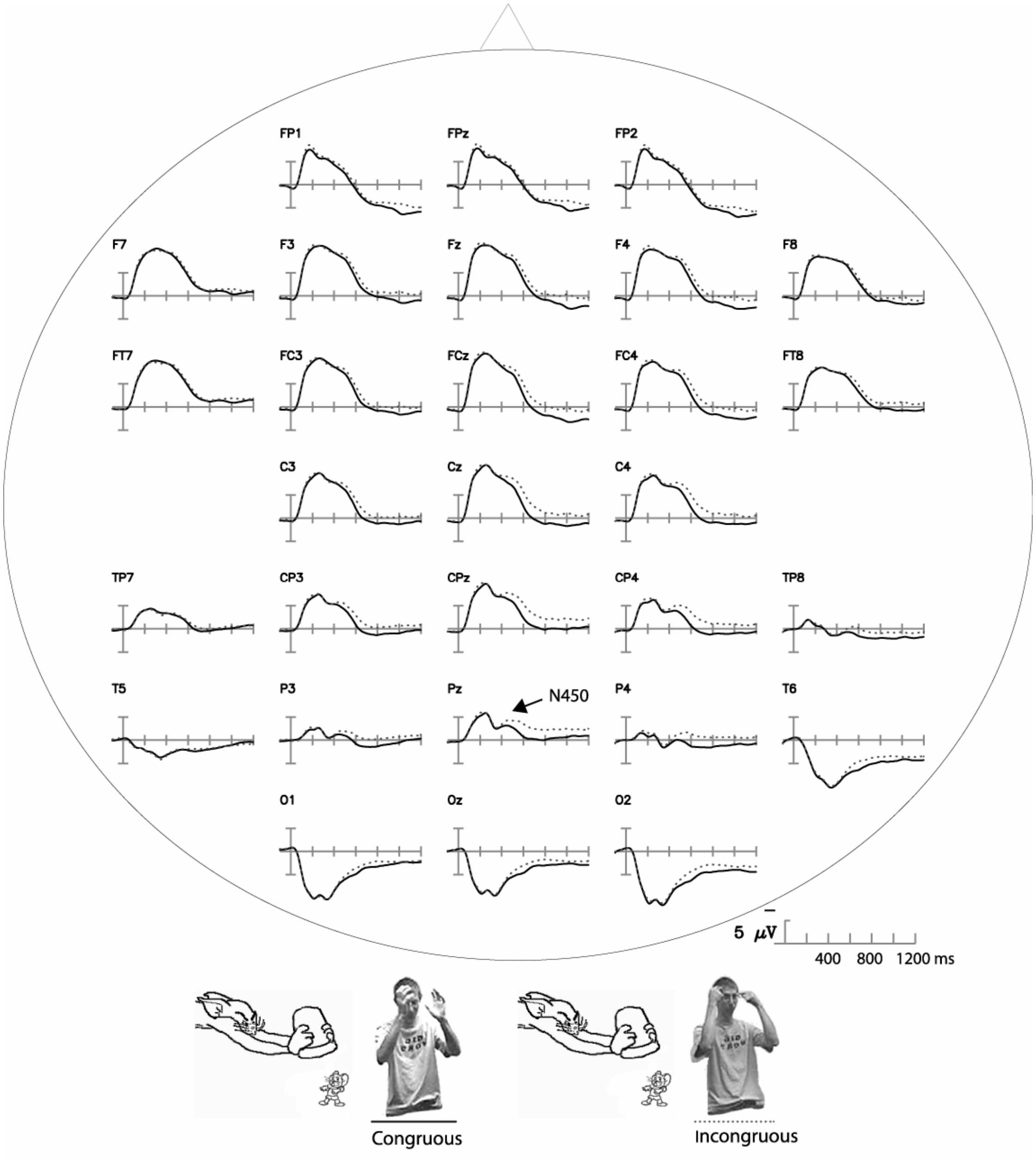


**Figure 2.4.** Experiment 2: Grand averaged ERPs time-locked to the onset of congruous and incongruous gestures and extending for 1200 ms. Time zero corresponds with the onset of gesture clips.

**Table 2.3.** 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 |

### *Median Split: Highly versus Moderately Congruent Trials*

Gesture congruency effects in highly and moderately related trials can be seen in Figure 2.5. Measured from 300 to 400 ms, no effects were observed, all F’s<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 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, F’s < 1, n.s.

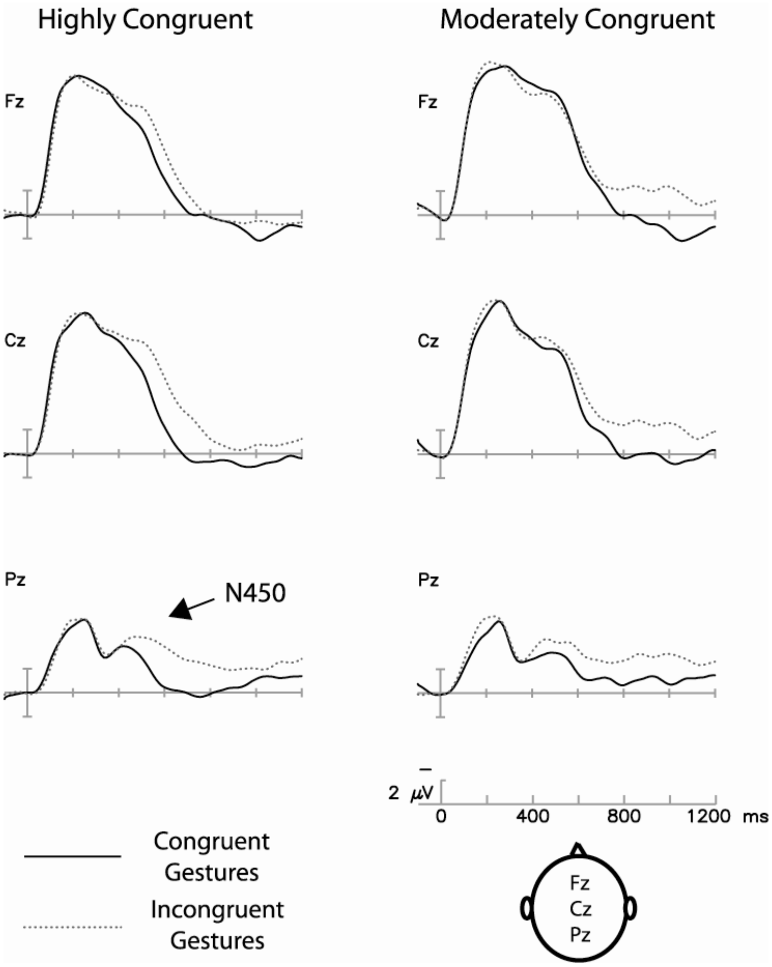


**Figure 2.5.** Experiment 2: Grand averaged ERPs time-locked to the onset of highly and moderately congruent gestures. N450 effects are larger for highly related items; late congruency effects are more prominent for moderately related ones.

From 600 ms post-stimulus 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 2.5).

***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 2.4). 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 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.

**Table 2.4.** 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) |

### *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 2.5).

**Table 2.5.** Mean Accuracy in Classifying Probe Words

| | Congruent | | Incongruent | |
|---|---|---|---|---|
| Related | .95 | (.01 SE) | .89 | (.02 SE) |
| Unrelated | .94 | (.02 SE) | .91 | (.03 SE) |

### *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 2.6).

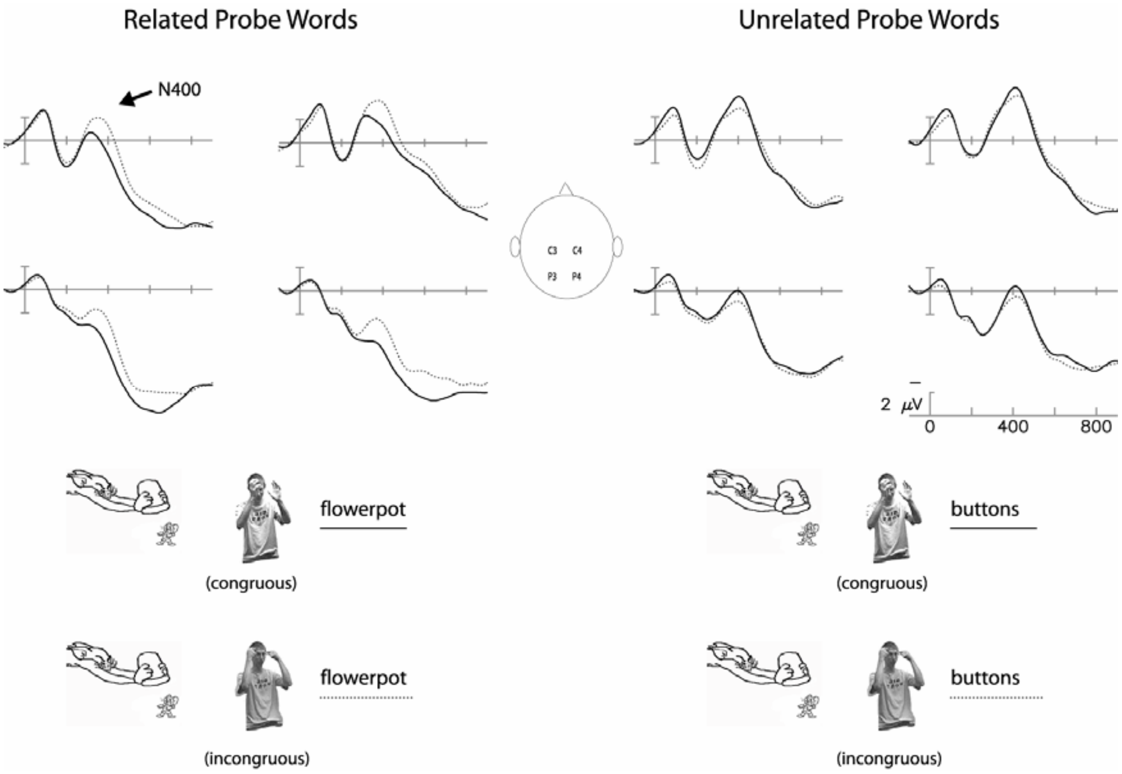


**Figure 2.6.** Experiment 2: N400 elicited by related and unrelated words following congruous and incongruous contexts.

Between 300 and 500 ms post-stimulus, 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, F's<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 post-stimulus, the presentation of incongruous gestures again resulted in more negative ERPs by 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. By 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 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 post-stimulus). Early congruency effects in Experiment 1 were likely due to the gesture classification task. In Experiment 2, by 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 2.6). 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

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 post-stimulus (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 2.7), their shared sensitivity to contextual congruity may indicate membership in a family of inter-related neural processes subserving contextual integration in different modalities.

One concern raised by this paradigm is the possibility that verbal information attained through lip-reading, 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° visual angle, minimizing the discernibility of lip movement and other non-auditory information deriving from the physical production of speech. Secondly, a recent (unpublished) study in our laboratory revealed that similar N450 effects are elicited by static gesture "freeze frames" –

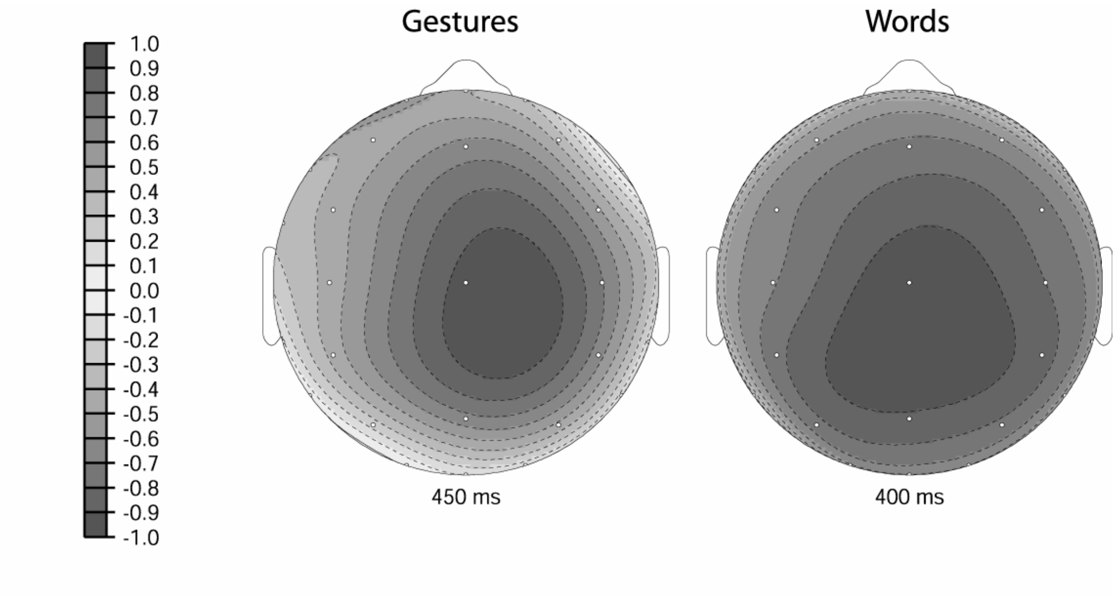


**Figure 2.7.** Experiment 2: Isovoltage maps show the distribution of gesture congruency effects at 450 ms post-onset (Incongruent minus Congruent), and word relatedness effects at 400 ms post-onset (Unrelated minus Related). Voltage in microvolts is represented on the scale bar. Values were normalized by dividing each data point for each stimulus type by the absolute value of the maximal data point for each stimulus type.

which do not represent any dynamic speech production information – extracted from the video streams used as stimuli in the present experiments (Wu, submitted).

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 2.8). 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 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 2.5 and 2.8). In some cases, the speaker's gestures depict actions that were not actually shown in the cartoon, but that could be inferred. For

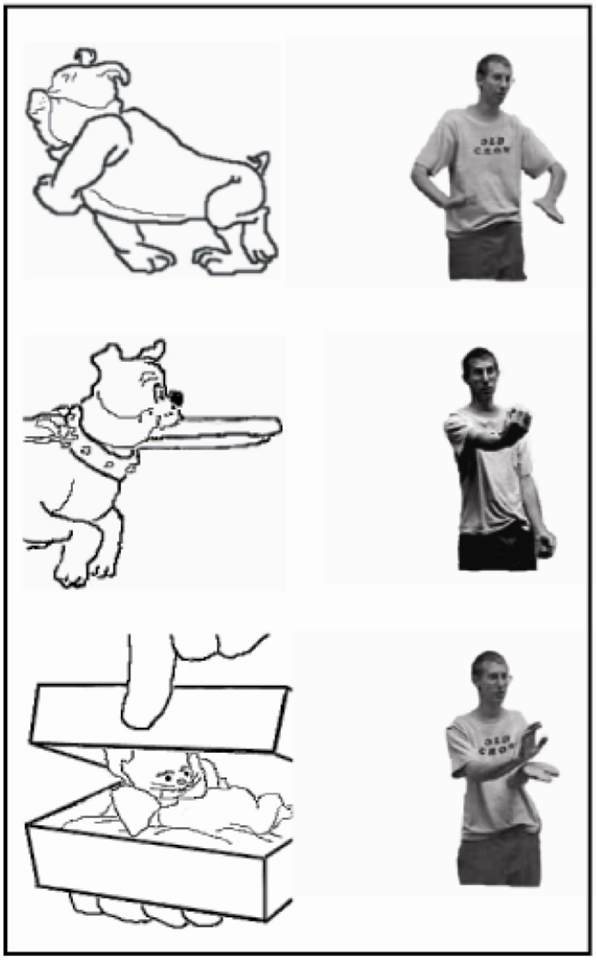


**Figure 2.8.** Examples of gestures used in Experiment 1 and 2. The speaker is portraying the dog's extended jaw, the opening box, and the dog's front legs.

example, in one trial, the speaker depicted how Elmer Fudd opens a large box with a crow-bar like instrument, though the cartoon clip only shows him lifting aside the lid. In other cases, the cartoon characters' actions often involved rapid, wide-ranging motions (e.g., leaping, diving, flying), that differed from the gestural enactments.

Finally, actions that occur on the left side of space in the cartoon are 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 2.1, 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 co-ordination 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 pre-activated 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.

## Acknowledgement

The authors would like to thank Marta Kutas, and three anonymous reviewers for suggestions on an earlier version of this 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.

Chapter 2, in full, has been published in *Psychophysiology* 2005, Wu, Ying Choon; Coulson, Seana, Blackwell Press, 2005. The dissertation author was the primary investigator and author of this paper.

## Works Cited

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*(2), 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*(6), 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), 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), 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: 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: Eletrophysiological 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*(2), 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), 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*(10), 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*(2), 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: 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*(1), 243-260.

Kendon, A. (2000). Language and gesture: unity or duality? In D. McNeill (Ed.), *Language and Gesture* (pp. 47-62). UK: Cambridge U 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*(5), 743-754.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12).

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: Chicago University 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*(3), 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*(1), 160-164.

Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Review of Neurobiology, 14*(3-4), 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*(6), 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*(3), 363-375.

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

# CHAPTER 3: LEXICAL PRIMING BY GESTURES

During discourse, speakers use hand and body movements to depict conceptual content salient to their talk. Individuals have traced an oval shape in the air while describing a platter, or demonstrated running legs by wiggling two fingers (McNeill, 1992; Wu & Coulson, 2005). Movements such as these, known as iconic gestures (McNeill, 1992), are proposed to affect listener comprehension. A number of behavioral studies have demonstrated listener sensitivity to information conveyed in gestures (Alibali et al., 1997; Cassell et al., 1999; Goldin-Meadow & Sandhofer, 1999), as well as improved comprehension of spoken discourse when speakers' gestures are visible (Beattie & Shovelton, 1999b, 2002; Rogers, 1978; Valenzeno et al., 2003). Measuring event-related potentials (ERPs), researchers have also shown differences in brain activity elicited by words presented with congruent relative to incongruent gestures, or no gesture (Kelly, Kravitz, & Hopkins, 2001). These findings suggest that iconic gestures are analyzed for meaning, and can produce measurable effects on observer comprehension.

Recent research has investigated commonalities in semantic processes mediating the comprehension of iconic gestures and the comprehension of more conventional visual representations such as pictures. Picture probes, for example, have been shown to elicit more negative ERPs around 300 ms (N300) and 400 ms (N400) post-stimulus when they are preceded by unrelated picture primes relative to related ones (Barrett & Rugg, 1990; Ganis et al., 1996). The N400 relatedness effect elicited by pictures is similar to the "classic" N400 elicited by words. Originally discovered in response to sentence final words, the lexical N400 was described as a negative-going deflection of the ERP waveform peaking between 300 and 500 ms post-stimulus, with an enhanced amplitude for incongruous items relative to congruous ones (Kutas & Hillyard, 1980a, 1984). Subsequently, similar effects were obtained in congruency manipulations involving a number of different stimulus types, including written and spoken word pairs (Holcomb & Neville, 1990), photographs (W. B. McPherson & P. J. Holcomb, 1999), and videotaped actions (Sitnikova et al., 2003). Further, ASL hand signs (Neville et al., 1997), which recruit more bilateral cortical resources than spoken language (Bavelier et al., 1998), and emblematic gestures (e.g., "thumbs up") (Gunter & Bach, 2004) have also elicited N400-like activity. These findings suggest that the N400 class of negativities, while probably resulting from overlapping, but non-identical neural generators, comprises a brain response triggered by meaningful stimuli. Because the amplitude of the N400 is inversely correlated with the degree to which an item is expected in its context (Kutas & Hillyard, 1984; Van Petten & Kutas, 1990), this brainwave component is generally thought to index the integration of incoming semantic information into a higher order mental model.

To test for N400-like responses to gestures, we recorded ERPs as participants watched spontaneously produced iconic gestures preceded by either congruous or incongruous cartoon contexts (Wu, 2005; Wu & Coulson, 2005). Relative to congruous trials, incongruous gestures elicited more negative ERPs between 450 and 600 ms post-stimulus (gesture N450). This effect displayed a similar time-course and polarity to the N400 family of negativities, as well as similar eliciting conditions. These data suggest that like words and pictures, iconic gestures also engage meaning-based representations that are integrated with other contextually active information.

Further support for this view was uncovered by extracting static freeze-frames from gesture videos, and pairing them with congruent or incongruent cartoon contexts (Wu, 2005). In addition to exhibiting enhanced N450, incongruent freeze-frames also elicited more negative ERPs between 300 and 400 ms (N300). Like the N400, the N300 is sensitive to manipulations of relatedness between images and prior context. However, because the N300 has only been observed in response to pictures and photos, it is thought to reflect the activation of image-specific conceptual representations. The finding that incongruent static gestures elicited larger N300 suggests that understanding these items' semiotic features (e.g. hand shape, location, orientation) recruits similar comprehension processes as well. It is possible that dynamic gestures did not yield N300 effects because processes indexed by the N300 might become activated slightly later in response to moving, visually complex stimuli than they would in response to static objects, overlapping with processes indexed by the N450.

The present study investigates whether iconic gestures activate meaning-based representations in the absence of supporting linguistic context. Because iconic gestures are not part of a conventionalized symbolic system, it has been argued that the their meaning is determined largely by speech accompanying them (Krauss et al., 1991). However, given evidence that understanding gestures engages semantic processes analogous to those recruited by pictures, people may be capable of integrating the semiotic features of gestures with stored knowledge about their referents even in the absence of contextual support.

To test this hypothesis, we recorded ERPs as healthy adults watched spontaneously produced iconic gestures followed by probe words. In Experiment 1, participants classified probes as related or unrelated; in Experiment 2, they attended to stimuli and completed a test of incidental probe recognition afterwards. If gestures activate stored knowledge about the phenomena they depict, evidence of word priming is expected in the form of reduced response latencies and reduced amplitude of the N400 elicited by related relative to unrelated probe words.

## EXPERIMENT 1

### Method

#### *Participants*

Sixteen volunteers were compensated for participation. 4 individuals were excluded due to excessive artifacts in the EEG. Data from the 12 remaining volunteers (5 women, 7 men) were included in the final analysis. All were healthy, English speaking adults with no history of neurological impairment.

#### *Materials and Procedure*

160 gesture video clips were paired with related or unrelated probe words. Gesture clips were constructed by videotaping a naive individual as he described segments of cartoons (e.g., Tom & Jerry). He was told the video would be utilized in a memory experiment, and instructed to describe each clip in as much detail as

possible. Iconic gestures digitized for experimental use typically either re-enacted actions performed in the cartoon (e.g., turning a doorknob), or depicted salient features of objects or events (e.g., the path of a careening rock).

12 volunteers (none of whom participated in the ERP experiment) rated the degree of relatedness between probe words and counterpart gestures on a five point scale. Related words were consistently rated as more related than unrelated ones, yielding a mean rating of 3.34 (0.2 SE) as compared to 1.55 (0.13 SE), $F(1,11) = 255$, $p < .0001$.

Four lists were constructed, each containing 40 related gesture-word pairs, 40 unrelated pairs, and eighty unrelated fillers. No gesture clip or word was repeated on any list, but across lists, each word appeared once as a related and once as an unrelated stimulus. Trials began with a fixation cross, presented in the center of a 17 in. color monitor. Gesture clips were presented at a rate of 48 ms per frame (48 frames total for each trial). One second after gesture offset, a probe word appeared for 1 s. This relatively long ISI allowed participants to refocus their gaze on the center of the screen. All frames were centered on a black background, subtending approximately 11° of visual angle.

Participants were told they would view soundless video clips showing a man describing cartoons, followed by probe words. They were asked to press either YES or NO on a button box as soon as they felt confident that the word either matched or did not match the preceding video. Because it was unclear whether decontextualized gestures contained enough information for the relatedness judgment task, participants were told that they might find a large proportion of probes unrelated. Response hand was counterbalanced across subjects. Four additional trials comprised a practice block.

### EEG Recording

The electroencephalogram was recorded using tin electrodes at 29 standard International 10-20 sites (Nuwer et al., 1999) (see (Wu & Coulson, 2005) for further explanation of electrode placement). Electrodes were also placed on the right mastoid for off-line re-referencing, and below the right eye and at the outer canthi for monitoring blinks and eye movements. All electrodes were referenced online to the left mastoid, and impedences maintained below 5 kΩ. EEG was amplified with an SA Instrumentation isolated bioelectric amplifier and digitized on-line at 250 Hz (band pass, 0.01 to 40 Hz).

Off-line re-referencing to averages of the right and left mastoids was performed after artifact removal. ERPs were time-locked to the onset of probe words, spanning a window from 100 ms before stimulus onset to 920 ms after. Only trials accurately categorized by participants were included in the averages. The mean artifact rejection rate was 17% (SD = 10%) for related trials and 18% (SD = 14%) for unrelated trials. On average, related bins contained 24 trials (SD = 5), and unrelated bins contained 27 trials (SD = 7).

### Data Analysis

Behavioral data were assessed with repeated-measures ANOVA. With subjects as a random variable, response latencies were trimmed within two standard deviations of each subject's mean latency for both conditions (4% of related trials and 5% of unrelated trials lost due to trimming); an additional analysis with items ($F_2$) as the random variable was also conducted.

### *EEG Analysis*

Mean amplitude and peak latencies of probe words were measured from 300 to 500 ms, and 500 to 900 ms post-stimulus onset. Measurements underwent repeated-measures ANOVA with the factors of Word Relatedness and Electrode Site. Scalp distribution of ERP effects was investigated as in (Wu, 2005). P-values were subjected to Geisser-Greenhouse correction (Geisser & Greenhouse, 1959) where appropriate, though original degrees of freedom are reported.

## Results and Discussion

### *Behavioral Responses*

Participants accurately classified 71% of related words (SE = 2%) and 85% of unrelated words (SE = 3%). Greater accuracy for unrelated words, $F(1,11) = 5.4$, $p < .04$, reflects participants' bias toward the unrelated response. The mean classification time for related items was 967.2 ms (SE = 51), and 1024 ms (SE = 77) for unrelated items. No effect of Word Relatedness was observed in either the subjects, $F = 2$, n.s., or items analysis, $F<1.5$, n.s.

### *ERPs*

ERPs to related and unrelated probe words are shown in Figure 3.1. Between 300 and 500 ms post-stimulus, the amplitude of the negative component (N400) was considerably more negative in response to unrelated items, as indicated by a main effect of Word Relatedness (see Table 3.1.) The N400 effect was maximal over

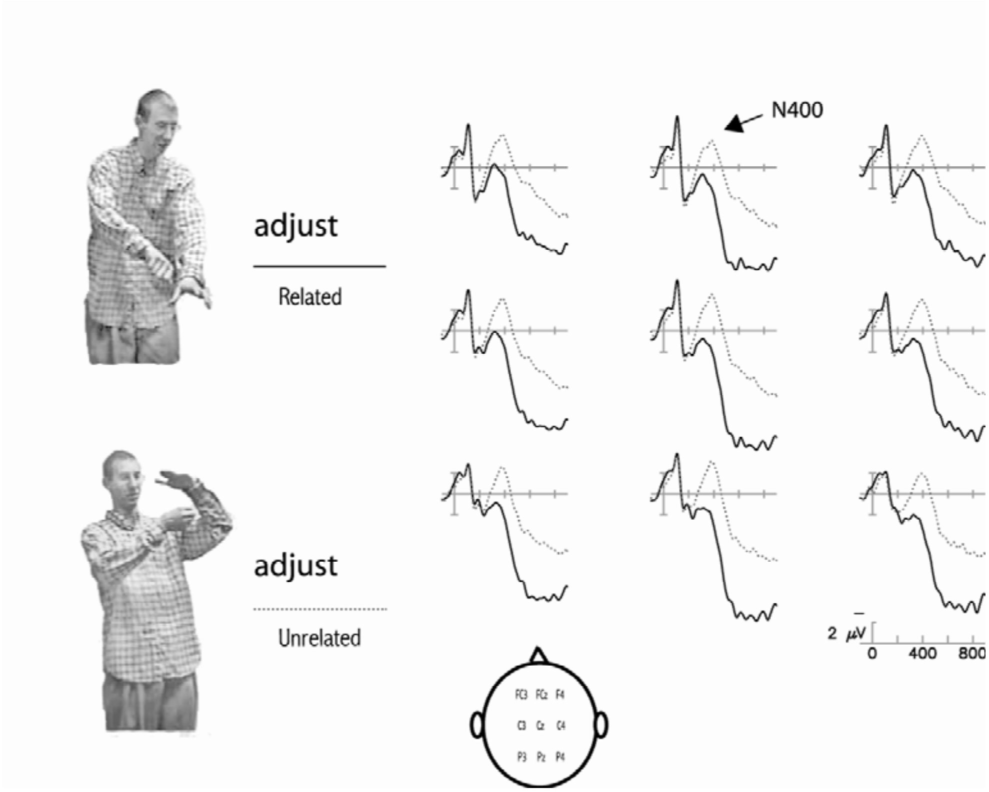


**Figure 3.1.** N400 elicited by related and unrelated words following spontaneously produced iconic co-speech gestures in Experiment 1. Data have been digitally filtered to remove frequencies greater than 15 Hz.

central and centro-parietal midline electrode sites (Cz and CPz), and was larger over posterior right hemisphere lateral electrode sites relative to corresponding ones over the left hemisphere (see Figure 3.3). From 500 ms to the end of the epoch (900 ms), unrelated words continued to elicit reliably more negative ERPs with a similar centro-parietal midline maximum.

**Table 3.1.** Experiment 1: Analyses of Mean Amplitude of ERPs Elicited by Probe Words

| Time Interval (ms) | Relatedness | | | Relatedness x Posteriority | | |
|---|---|---|---|---|---|---|
| | Midline | Medial | Lateral | Midline | Medial | Lateral |
| | F(1,11) | F(1,11) | F(1,11) | F(6,66) | F(6,66) | F(3,33) |
| 300 – 500 | 24.7*** | 26*** | 27.2*** | 2.9 | 1.1 | 1.1 |
| 500 – 900 | 35.8*** | 29.7*** | 15.5** | 7.6** | 4.4* | 1.5 |

| Time Interval (ms) | Relatedness x Hemisphere | | Relatedness x Hemisphere x Posteriority | |
|---|---|---|---|---|
| | Medial | Lateral | Medial | Lateral |
| | F(1,11) | F(1,11) | F(6,66) | F(3,33) |
| 300 – 500 | 1.9 | 4.5† | 1.6 | 7.6** |
| 500 – 900 | 2.5 | 2.3 | 0.4 | 2.5 |

*p<.05
**p<.005
***p<.0005
†p=.06

Words related to preceding gestures elicited less N400 than unrelated ones, in keeping with the view that gestures activate information that facilitates lexical integration. However, ERPs to related words are also more positive-going than their unrelated counterparts in the latter portion of the epoch. This positivity is likely a member of the P300 family of potentials, which tend to be enhanced in response to targets presented in binary decision paradigms (see (Kok, 2001) for review). It is unclear to what degree the observed relatedness effect was driven by the binary word classification task, and to what degree it genuinely reflects facilitation of processes sensitive to the meaningful content of gestures.

## EXPERIMENT 2

To eliminate ERP effects attributable to the classification task, we conducted a second experiment. A new group of volunteers was presented with the same gesture-word pairs, but only instructed to attend to stimuli. Afterwards, participants received a surprise memory test, assessing words recognized from the experiment. This test provided an indirect measure of attention to probes. We predicted that if gestures aid word comprehension, unrelated words would elicit larger N400 than related ones.

### Methods

Fourteen volunteers participated in Experiment 2. Two individuals were excluded due to excessive artifacts. Twelve remaining volunteers (7 women, 5 men) were included in the final data set. Materials and presentation parameters were identical to Experiment 1, with the exception that filler trials were omitted, yielding a stimulus set of 80 items. Instructions were to attend to video clips and words. The incidental memory test administered afterwards contained all probe words presented during the experiment, along with 80 distractors. EEG recording and analysis were identical to Experiment 1. Critical bins contained an average of 34 trials (SD = 5). The mean artifact rejection rate was 14% (SD = 12%).

## Results and Discussion

### *Accuracy*

On average, participants accurately recognized 50% of probe words (SD = 17%): 27.5% of related words (SD = 9%) and 22.5% of unrelated ones (SD = 11%). This difference approached significance in a two-tailed matched pairs *t* test, t(11)= -1.9, p=.09).

### *ERPs*

As shown in Figure 3.2, unrelated words consistently elicited more negative ERPs than related ones between 300 and 500 ms, and 500 to 900 ms after stimulus (see Table 3.2). These outcomes demonstrate that even without explicit analysis, the visuo-spatial cues provided by gestures make the comprehension of related words easier.

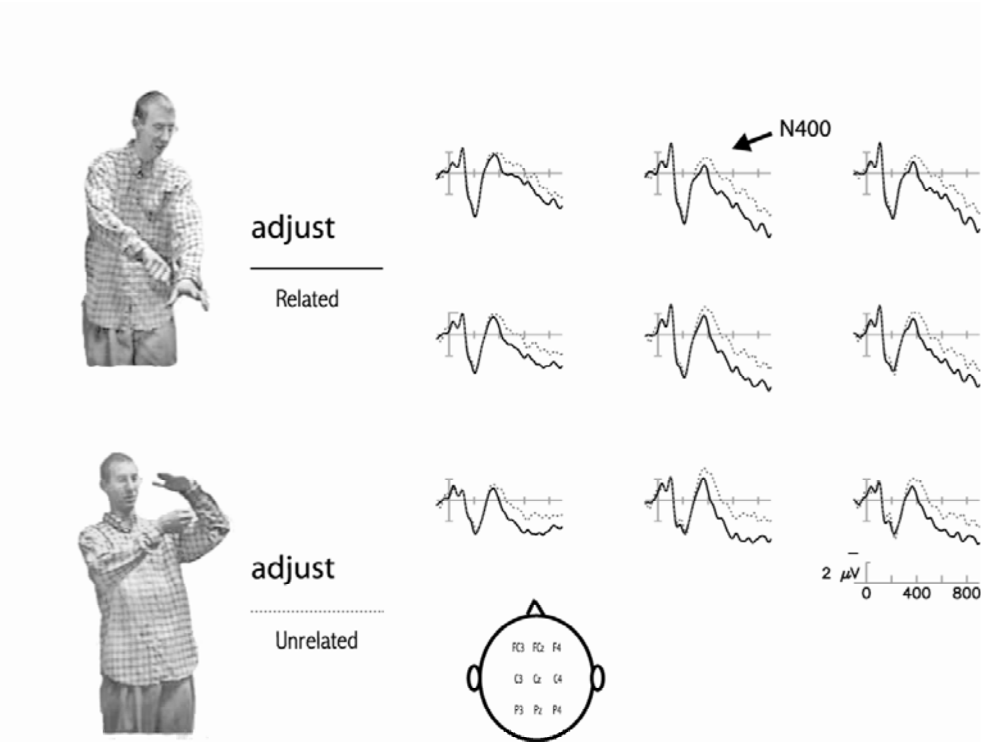


**Figure 3.2.** N400 elicited by related and unrelated words following spontaneously produced iconic co-speech gestures in Experiment 2.

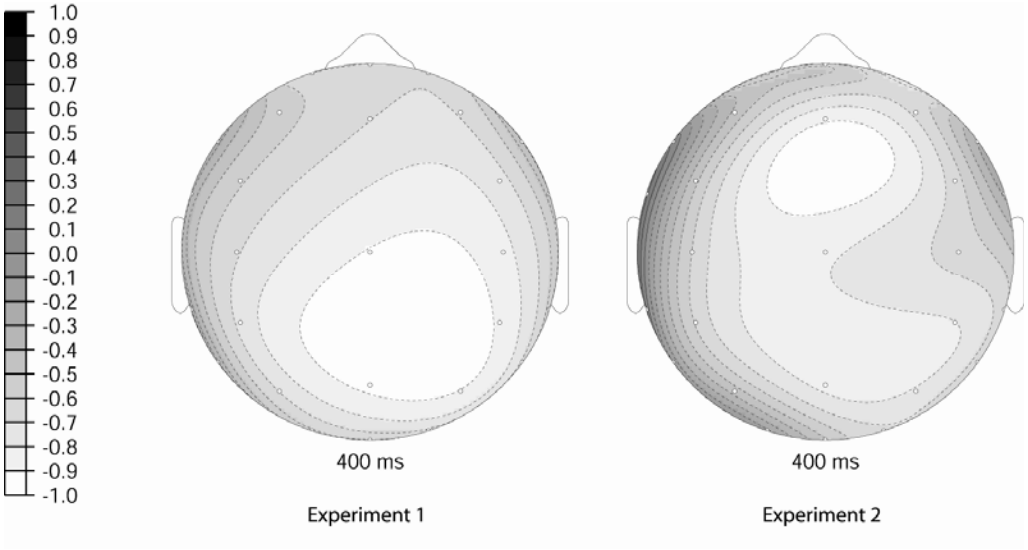


**Figure 3.3.** Topography of N400 effect in Experiment 1 and 2.

**Table 3.2.** Experiment 2: Analyses of Mean Amplitude of ERPs Elicited by Probe Words

| Time Interval (ms) | Relatedness | | | Relatedness x Posteriority | | |
|---|---|---|---|---|---|---|
| | Midline | Medial | Lateral | Midline | Medial | Lateral |
| | F(1,11) | F(1,11) | F(1,11) | F(6,66) | F(6,66) | F(3,33) |
| 300 – 500 | 6.7* | 5.8* | 2.9 | 0.8 | 0.4 | 0.4 |
| 500 – 900 | 5.6* | 5.2* | 3.5 | 1.0 | 0.5 | 1.1 |

| Time Interval (ms) | Relatedness x Hemisphere | | Relatedness x Hemisphere x Posteriority | |
|---|---|---|---|---|
| | Medial | Lateral | Medial | Lateral |
| | F(1,11) | F(1,11) | F(6,66) | F(3,33) |
| 300 – 500 | 2.7 | 6.3* | 1.5 | 1.3 |
| 500 – 900 | 1.2 | 3.1 | 1.2 | 0.6 |

*p<.05

## General Discussion

To assess priming by spontaneous, iconic gestures, we recorded ERPs as healthy adults watched soundless gesture clips followed by related and unrelated probe words in two experiments. In Experiment 1, participants judged the relatedness between gestures and probe words. When collapsed across lists, the same probes appeared in both related and unrelated trials; yet, related words were classified less accurately than unrelated ones (71% vs 85% for unrelated words), indicating that in the absence of supporting context, the intended meaning of gestures was apprehended with difficulty, though at an above chance rate.

Moreover, ERPs to related and unrelated probe words differed approximately 350 ms after stimulus onset. This effect exhibited a centro-parietal, right-hemisphere lateralized distribution which is typical of the N400 effects reported in other studies of visually presented words (Kutas, Van Petten, & Besson, 1988). Unrelated words elicited larger amplitude N400 than related ones. In Experiment 2, participants were instructed simply to attend to gestures and words, and recognition of incidentally encoded words was measured after the EEG recording session. As expected, unrelated words elicited larger N400 than related ones. Though more broadly distributed than in Experiment 1, this effect also appeared larger over right hemisphere electrode sites. Differences in topography of N400 effects observed in the two experiments may be attributed to different demands of the categorization and recognition tasks.

Outcomes from Experiment 2 demonstrate that even when no explicit relatedness judgment is required, gestures nevertheless benefit the processing of words. This result is consistent with numerous other studies reporting N400 context effects regardless of whether participants are asked to make explicit semantic judgments. N400 effects have been observed for targets in masked priming paradigms where participants are unable to report the prime (Deacon, Hewitt, Yang, & Nagata, 2000), and in attentional blink paradigms where participants are unable to report the target (Luck, Vogel, & Shapiro, 1996). Indeed, with sufficient contextual support, N400 effects begin before word recognition is complete (Van Petten, Coulson, Rubin, Plante, & Parks, 1999).

Response latencies also reflected sensitivity to word relatedness. On average, individuals classified related words 57 ms faster than unrelated ones, though this effect did not exceed the threshold of conventional significance. Non-significant effects here may be due to the relatively small sample size.

In the present study, gestures may have activated semantic or associative information related to referents whose visuo-spatial properties were consistent with the gestures' semiotic features. We suggest that these activations were sufficient to reliably facilitate the integration processes indexed by the N400, but not those required for the relatedness judgment task. This pattern is similar to that reported in a study investigating semantic priming by words whose referents bear visuo-perceptual resemblances to the referent denoted by the target word (e.g. *button – coin*), in which an N400 effect was reported, but behavioral effects were non-significant (Kellenbach, Wijers, & Mulder, 2000). Similarly, Pecher and colleagues (1998) report that RTs revealed evidence of visuo-perceptual semantic priming only in cases where participants were induced to process perceptual properties associated with words, and to avoid relatedness-checking strategies. Thus, it is possible that some behavioral tasks may not be sensitive to the pre-activation of visuo-semantic features of words.

The iconic gestures studied here fall mid-way between pointing gestures (Kelly et al., 2001) and symbolic hand signs (Gunter & Bach, 2004) studied previously in terms of their similarity to linguistic systems such as English or American Sign Language. Like linguistic symbols, emblematic hand signs such as the "thumbs up" gesture are conventionalized and arbitrarily[2] related to their referents. In contrast, the link between iconic gestures and their referents is based on fairly abstract perceptual similarities. Moreover, while an essential property of language is displacement, or the capacity to communicate things that are not currently present, pointing gestures as

---

[2] Some analysts might argue that emblems exploit metaphoric iconicity; however, such mappings are even more abstract than those necessary for the interpretation of iconic gestures used in the present study.

studied in (Kelly et al., 2001) are meaningful because their referent is co-present. By contrast, the iconic gestures we used were presented in the absence of their referents. Further, unlike previous research on the brain response to gestures, our gesture stimuli were not artificially generated by actors, but emerged naturally from speakers in the course of authentic communicative situations.

What does the finding of lexical priming by iconic gestures reveal about comprehension in natural conversation, where talk and gesture unfold together across both visual and auditory modalities? Our findings suggest that iconic gestures activate meaning-based representations which are compatible with linguistic surface forms, in keeping with other studies showing cross-modal priming of words by pictures (Carr et al., 1982; Hines, 1993; Pratarelli, 1994; Vanderwart, 1984), and lexical priming on the basis of visual-perceptual similarities between the referents of two words (Kellenbach et al., 2000). Comprehension processes prompted by iconic gestures may not dramatically differ from those prompted by other contentful representations, such as words and pictures. Whether iconic gestures facilitate speech comprehension is a matter for further investigation.

This study enhances the current field of gesture research in two important ways. First, it investigates real time processes mediating multi-modal discourse comprehension, complementing previous behavioral research on the communicative value of iconic gestures. Second, it provides evidence that body movements which are neither conventionally meaningful or co-present with their referent can affect the processing of related words. Our finding that spontaneously produced iconic gestures prime related concepts suggests commonalities in the processing induced by iconic gestures with that prompted by entrenched hand signs (Gunter & Bach, 2004) as well as pointing gestures made in the presence of their referents (Kelly et al., 2001). The findings reported here indicate that the abstract interpretation of human biological movement can affect the comprehension of linguistic surface forms.

## Acknowledgment

The authors thank Thom Gunter and Aaron Newman for comments on this research. Y. W. was supported by Institute for Neural Computation NIH Training Grant #MH 2002-07 and Center for Research in Language NSF Training Grant DC 000-16.

Chapter 3, in full, is currently in press for publication of the material as it appears in *Psychonomic Bulletin and Review*, Wu, Ying Choon; Coulson, Seana, Psychonomic Society Publications, 2006. The dissertation author was the primary investigator and author of this paper.

## Works Cited

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.

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., et al. (1998). Hemispheric specialization for English and ASL: left invariance-right variability. *Neuroreport, 9*(7), 1537-1542.

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*(2), 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*(6), 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), 1-33.

Deacon, D., Hewitt, S., Yang, C. M., & Nagata, M. (2000). Event-related portential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research, 9*, 137-146.

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*(2), 89-106.

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

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

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

Hines, D. (1993). Effect of masked picture primes on semantic priming for easy- and difficult-to-name words. *Journal of General Psychology, 120*(2), 149-169.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281-312.

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research, 10*, 67-75.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2001). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 243-260.

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

Krauss, R. M., Morell-Samuels, P., & Colasante, C. (1991). Do Conversational Hand Gestures Communicate? *Journal of Personality and Social Psychology, 61*(5), 743-754.

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., Van Petten, C. K., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*(3), 218-233.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616-618.

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

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

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating american sign language: Effects of sensory experience and age of acquisition. *Brain and Language, 57*(3), 285-308.

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).

Pecher, D., Zeelenberg, R., & Raajimakers, J., G. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory & Language, 38*, 401-418.

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

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

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teacher's gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology 29*(187-204).

Van Petten, C. K., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394-417.

Van Petten, C. K., & Kutas, M. (1990). Interactions between sentence context and word frequency in even-related brain potentials. *Memory & Cognition, 18*(4), 380-393.

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

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

Wu, Y. C., & Coulson, S. (2005). Is that a meaningful gesture: Electrophysiological indices of gesture comprehension. *Psychophysiology*, *42*(6), 654-667.

## CHAPTER 4: SEMANTIC INTEGRATION OF GESTURES AND SPEECH

Co-speech gestures provide a channel for speakers to express additional information related to their communicative intent. While uttering, "It's actually a double door," for example, a speaker may indicate the shape of a Dutch rather than French style door with the configuration of his hands (see Figure 4.1). A number of behavioral studies suggest that gestures such as these play a beneficial role in communication. Listeners rely on speakers' gestures to disambiguate communicative intent in cases where understanding may be impeded – due to noise in the speech signal, for example (Rogers, 1978; Thompson & Massaro, 1986, 1994), or due to additional inferential processing engendered by indirect requests (Kelly, 2001; Kelly et al., 1999). Listeners also exhibit a more accurate understanding of instructions and narratives when the speaker's accompanying gestures are visible (Beattie & Shovelton, 1999b, 2002; Graham & Argyle, 1975; Morford & Goldin-Meadow, 1992; Singer & Goldin-Meadow, 2005; Valenzeno et al., 2003). However, see Krauss, Dushay, Chen, & Rauscher (1995) and Goldin-Meadow & Sandhofer (1999) and for an alternative view.

These findings suggest that some properties of gestures may activate semantic information related to the content of the talk in progress. However, little is known about the cognitive and neural processes mediating this remarkable feat of multi-modal integration. Given growing interest in the role of motor mirroring systems in action comprehension (Rizzolatti & Arbib, 1998), the study of gesture may provide

**Figure 4.1.** The speaker’s hand configuration while uttering, “It’s actually a double door.”

cognitive neuroscience with a further venue for understanding the relationship between sensori-motor and higher order conceptual processing.

It has been proposed by McNeill and others that during comprehension, speech and gesture are integrated into a common underlying conceptual representation. He writes, “…listeners, after a brief delay, cannot tell whether information came to them in gesture or in speech, the two having become unified”; (McNeill, 1998). In support of this idea, a number of studies have investigated the comprehension of discourse in which the speakers’ gestures express something different from their words (as in the Dutch door example). By assessing listeners’ own accounts of what they had understood (Alibali et al., 1997; Cassell et al., 1999; Goldin-Meadow et al., 1992; Kelly & Church, 1998), or their responses on questionnaires (Goldin-Meadow &

Sandhofer, 1999), it has been demonstrated that listeners are sensitive to information made available in both modalities.

The goal of the present study is to investigate how speech and gesture affect real-time interpretation processes. Previous behavioral research has demonstrated that information presented to listeners in gesture, but not directly in speech, is accessible in long-term memory. However, little is known about the encoding processes whereby gesture-based information enters memory systems. Further, semantic activations induced by co-speech gestures have only begun to be investigated. The present study addresses the cognitive and neural processes mediating speech-gesture integration.

Recent research involving event-related potentials (ERPs) has begun to shed light on this question. ERPs represent dynamic voltage fluctuations that derive from synaptically generated current flow on the cortical surface of the brain. Tiny signals detectable at the scalp (on the order of microvolts) are amplified and digitized, yielding a record of on-going brain activity in the form of an electroencephalogram (EEG). By averaging portions of EEG recorded in synchrony with the presentation of a specific class of stimuli, it is possible to draw inferences about cognitive processes engaged by that type of stimulus. Because scalp-recorded potentials typically reflect contributions from a number of different neural sources, it is necessary to average event-related responses across many trials in order to cancel out random noise introduced by background neural activity. The resulting ERP waveform can be analyzed as a series of positive- and negative-going deflections (commonly referred to as components) that are characterized by their amplitude, time course and a distribution across scalp electrode sites.

A component particularly relevant to semantic processing is the N400, which was discovered during early research on language processing (Kutas & Hillyard, 1980b). Kutas and Hillyard recorded ERPs to the last word of sentences that either ended congruously (as in [1]), or incongruously (as in [2]).

(1) I take my coffee with cream and sugar.
(2) I take my coffee with cream and dog.

By averaging the signal elicited by congruous and incongruous sentence completions, respectively, these investigators were able to reveal systematic differences in the brain's electrical response to these stimulus categories occurring approximately 400 ms after stimulus onset. Subsequent research has shown that N400 components are generated whenever stimulus events involve meaningful processing of the stimuli, and that its size is sensitive to fairly subtle differences in the processing difficulty of the words that elicit it. As such, many investigators have used the N400 component of the brain waves as a dependent variable in psycholinguistic experiments (see (Kutas, Federmeier, Coulson, King, & Muente, 2000) for review).

To investigate the effect of gestures on language comprehension, Kelly, Kravitz, & Hopkins (2004) recorded ERPs elicited by spoken words articulated in synchrony with gestures that were either congruent and incongruent with word meanings. Stimuli were constructed by video taping 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. In the matching condition, the actor's speech corresponded with both the object as well as the spatial dimension indicated in gesture. In the complementary condition, the speech token described a different dimension of the referent from that depicted by the gesture (e.g., *tall* uttered in accompaniment with a gesture indicating the thin diameter of the glass). In the mismatch condition, the speech token corresponded to one object while the gesture corresponded to the other. Finally, in the no gesture condition, speech was presented alone.

Results yielded early effects of gesture congruency (between 100 and 352 ms), with mismatching and complementary stimuli eliciting relative to other conditions larger P1 and P2 components, which reflect auditory sensory processing. N400-like effects were also observed at bilateral temporal electrode sites, with mismatch trials eliciting more negative ERPs than all other conditions around 450 ms post-stimulus. These findings suggest that gesture congruency affects both early sensory as well as higher order semantic processing of words.

Other studies have approached the neuro-cognitive underpinnings of gesture comprehension by measuring ERPs elicited by gestures themselves. Besides words, the N400 component has also been elicited by pictorial stimuli. For example, line drawings of objects elicit a larger N400 when they are preceded by an unrelated word than by a related word (Praterelli, 1994). Likewise, both line drawings and photographs of objects elicited a larger N400 when preceded by pictures of semantically unrelated than related objects (P. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. Holcomb, 1999).

To test whether gestures engage semantic processes similar to those engaged by pictures and words, Gunter and Bach compared ERPs elicited by meaningless hand configurations and emblematic gestures (e.g., thumbs up, OK sign; (Gunter & Bach, 2004). They reported enhanced N400 to meaningless compared to meaningful stimuli. Similarly, spontaneous iconic gestures produced in the course of conversation elicited more negative ERPs between 400 and 600 ms post-stimulus (gesture N450) when preceded by incongruent contexts relative to congruent ones (Wu, 2005; Wu & Coulson, 2005). Due to its functional characterization, the gesture N450 was interpreted as a member of the N400 family of negativities. These findings suggest that in spite of their sometimes idiosyncratic, schematic, and dynamic qualities, gestures are subject to semantic processes.

Rather than examining the effect of gestures on speech processing, as in Kelly et al (2004), or the effect of context on gesture processing, as in Wu & Coulson (2005), the goal of the present study was to investigate the respective contributions of speech and gesture to discourse comprehension. In keeping with McNeill (1998), we hypothesize that gestures activate semantic information related to the content of the talk in progress, allowing listeners to form a more robust representation of the speaker's intended meaning.

Research on the role of perceptual simulation in sentence comprehension offers experimental precedence for the view that language comprehension involves the activation of visuo-spatial properties even when they are not verbally expressed. Healthy adults have been presented, for example, with sentences followed by pictures showing an object configured in manner that was either congruent or incongruent with a given shape or orientation implied by within the sentence. An eagle with outstretched wings or folded ones followed the sentence, *The ranger saw the eagle in the sky.* Similarly, either a horizontal or an upright pencil followed *John put the pencil in the drawer*. Pictures that were congruent with implied spatial information were named or classified more quickly than incongruent ones (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). Studies such as these demonstrate that even in the absence of explicit encoding, visuo-spatial features including size, shape, and orientation are important for language comprehenders. When sufficient information is available from sentences to draw inferences about these features, they become active in the listeners' conceptual models. Iconic gestures, we propose, may also serve as an analogue resource for prompting such activations.

To test this proposal, we recorded ERPs as healthy adults viewed short segments of spontaneously produced discourse involving both descriptive speech and gestures. Our stimuli involved primarily iconic gestures because they typically express visuo-spatial relations that do not easily lend themselves to linguistic encoding (Emmorey & Casey, 2001). In order to assess semantic activations attributable to gestures, we measured the brain response to probe images that either did or did not reflect the spatial information conveyed by gestures in each discourse segment.

Aside from our utilization of a real-time measure of neural processing, the present study differs from previous research on speech-gesture integration in at least two important ways. First, the discourse segments used in this research were derived from spontaneously produced conversation. Studies involving rehearsed speech-gesture mismatches created by actors (such as Cassell et al (1999) or Kelly et al

(2004)) may not be reflective of comprehension processes engaged during every day conversation. Secondly, the present study differs from much existing research using spontaneous gesture stimuli in that the speaker does not make reference to co-present material structure. This feature enabled us to test visuo-semantic activations prompted exclusively by gestures.

Each discourse prime was followed by one of four types of picture probes (see Figure 4.2). Cross-modal related items were congruent with both the verbal and gestural component of the speaker's description. Speech-only related items were congruent only with descriptive features expressed through speech. A third logical "gesture-only related" condition – where pictures would be congruent with gestures, but not speech – was not included. Instead, two types of unrelated stimuli were created by pairing cross-modal and speech-only related items with other discourse primes, allowing each related picture to be compared with itself as an unrelated item.

We expected differences in the processing of cross-modal and speech-only probes to be reflected in two ERP-components sensitive to semantic relatedness between images and prior context. Previous studies comparing brain activity elicited by the second member of related and unrelated picture pairs have reported in addition to the N400, an earlier negative going deflection of the ERP waveform peaking around 300 (N300) ms after stimulus onset. The amplitudes of the N300 and N400 are both larger (more negative) in response to unrelated items (Barrett & Rugg, 1990; P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. Holcomb, 1999). This pattern of results has also been observed in experimental paradigms involving cross-modal priming between words and pictures (Hamm et al., 2002), pictorial

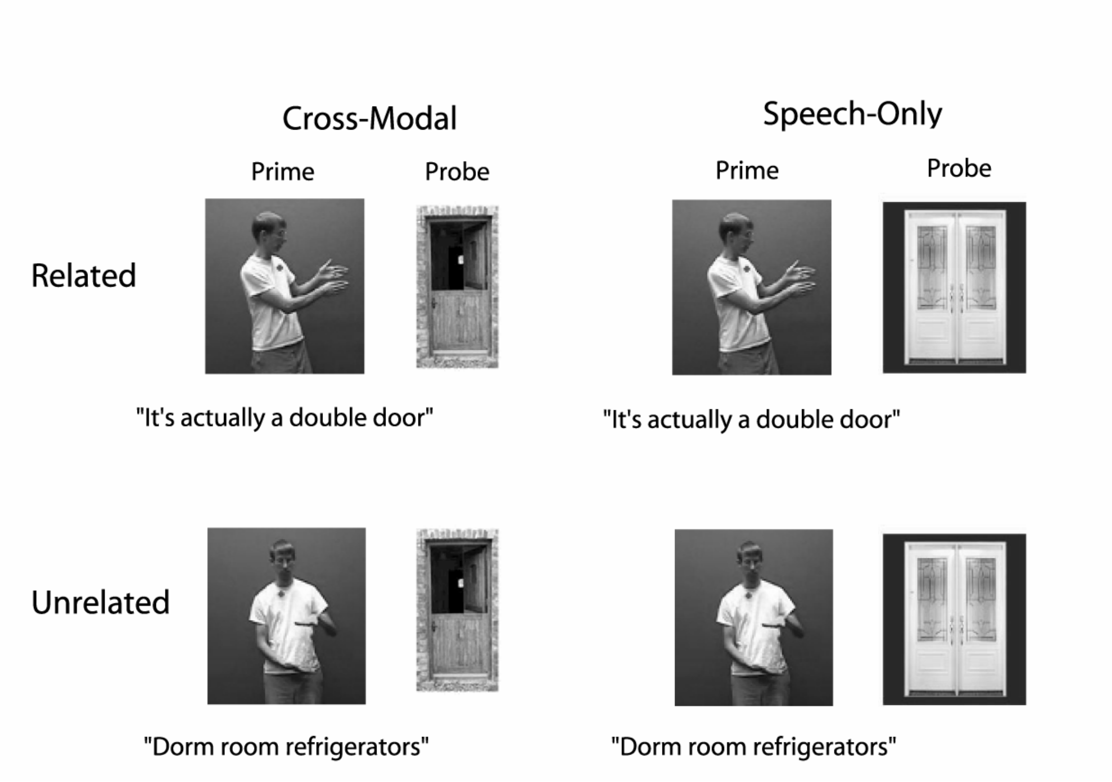


**Figure 4.2.** Experimental design. Cross-modal and speech-only picture probes occurred in related and unrelated trials.

completions of written sentences (Federmeier & Kutas, 2001, 2002), and complex scenes (West & Holcomb, 2002). When viewed in concert, these findings suggest that the N300 and N400 index the integration of incoming visual semantic input with recently activated stored knowledge.

Although both the N300 and N400 are modulated by contextual congruity, they are thought to reflect slightly different aspects of image comprehension. Because the N300 has only been observed in response to photographs or pictures, it has been proposed to index image-specific semantic processes. By contrast, the N400 has been found in studies involving a broad range of meaningful representations, including words (see Coulson & Van Petten (2002) for review), ASL hand signs (Neville et al., 1997), gestures (Gunter & Bach, 2004; Wu, 2005; Wu & Coulson, 2005), and action videos (Sitnikova et al., 2003).

The N400 is also functionally distinct from the N300 in its sensitivity to degrees of semantic fit. Whereas the amplitude of the N300 reflects differentiation only between unrelated and related trials, the amplitude of the N400 increases in a graded fashion in response to highly related, moderately related, and unrelated items (W. B. McPherson & P. J. Holcomb, 1999). On the basis of these findings, it has been proposed that the N300 reflects image-specific semantic processes, whereas the N400 reflects more general semantic integration processes.

If the semantic activations engendered by speech and gesture jointly contribute to the ongoing formation of discourse-level representations, then pictures that relate to information made available in both modalities should be easier to process than pictures that relate only to the speech, as indexed by the amplitude of the N300 and N400 ERP components. In general, we expect unrelated probes to elicit more negative ERPs than related probes. If co-speech gestures activate image-based representations, N300 relatedness effects should be more pronounced for cross-modal than for speech-only probes. Further, if general semantic integration processes indexed by the N400 are sensitive to information expressed both through speech and gesture, then N400 relatedness effects should also be larger for cross-modal than speech-only probes.

## Method

### ***Participants***

Sixteen volunteers (6 women and 10 men) were paid $16 or received academic course credit for their participation. All participants were healthy, fluent English speakers with no history of neurological impairment (mean age = 20, SD = 2). Fifteen individuals were right handed, and one was ambidextrous. The Edinburgh Inventory (Oldfield, 1971), revealed a mean laterality quotient of 0.73, in keeping with participants' self-reported right handed bias.

### ***Materials***

Stimuli included 168 video clips in which a speaker described a common object or event. In each case, the speaker's talk and gesture conveyed complementary, but not identical information. In one trial, for instance, he says, "Two throw pillows," while indicating in gesture that they are located at opposite ends of a couch. In another trial, he describes a hammer by saying, "the handle...the handle is wooden," while showing the object's horizontal orientation with his hands.

Video clips were followed by either related or unrelated picture probes. Related probes either agreed with both the speaker's speech and his gestures (cross-modal related), or they agreed with his speech alone (speech-only related). In the case of the throw pillows, the cross-modal related item was a sofa with pillows on either end, whereas in the speech-only related item, the probe depicted a sofa with adjacent throw pillows. In the case of the wooden hammer handle, the same hammer was shown at both a horizontal (cross-modal related) and vertical (speech-only related) orientation. Unrelated trials were constructed by pairing the same picture probes with different discourse primes, yielding a 2 × 2 factorial design with two levels of relatedness (related, unrelated) and two levels of stimulus type (cross-modal, speech-only), with each stimulus serving as its own control (see Figure 4.2).

Video clips were constructed by filming a naive individual as he described everyday activities, as well as photographs of common objects and scenes to an off-camera interlocuter. Six recording sessions took place. He was told that the video footage would be utilized in the construction of stimuli for a subsequent memory experiment; no mention of gestures was made. Experimental materials involved instances in which the speaker's spontaneous gestures conveyed information over and above that in his speech. These instances were captured and digitized into short video clips ranging in length from 2.6 to 7.6 s. Gestures either re-enacted elements of everyday actions (turning a knob, shaking out clothes, making the bed), or depicted affordances or spatial features of objects and scenes (the shape of a vase, the handles on a canvas bag, the location of a door). A total of 168 experimental clips were constructed, along with 7 filler and 2 practice clips.

Picture probes were constructed by collecting digital photographs from internet databases. Both picture probes and video frames were centered on a black background and subtended approximately 8.3° visual angle horizontally and 6.2° vertically. Within the videos, the speaker himself subtended approximately 5° to 6° vertically and 3° to 4.5° horizontally, and primarily the head, arms, and upper torso of the speaker were shown. Within the picture probes, depicted items subtended between 2.4° and 5° and were surrounded by a white background frame.

A normative study was conducted to evaluate the identifiability of probe materials. Twelve individuals viewed probes and either named the depicted objects or indicated that they could not identify them. On average, both probe types elicited names in 96% (SD =3 .6%) of responses, indicating that the two probe types were well balanced in terms of identifiability.

To evaluate how consistently probes were named, we defined the *common name* for each picture as the word that occurred most consistently in our informants' responses. For example, when presented with the cross-modal probe depicting a Dutch door (see Figure 4.2), 3 individuals responded with the word, "door." Other responses included "pub door," "sectional house door," and "wooden door to brick house." We selected *door* as the common name for this probe. After identifying the common name for each picture, we calculated the proportion of hits for that name out of the total number of responses received. On average, cross-modal pictures elicited their most common names in approximately 72% of responses (SD = 25%), and speech-only pictures, in approximately 75% (SD = 27%) of responses. This outcome suggests that the content of both types of picture probes was identified at a consistent rate well above chance. Further, on 60% of trials, the most common name elicited by cross-modal and speech-only probes was identical. Finally, we used Latent Semantic Analysis (Landauer & Dumais, 1997) to assess the degree of semantic similarity between the names elicited by each probe type. On a scale of 1 to -1 (with 1 indicating shared identity), the mean similarity between names was .69, (SD = .4).

To determine whether names elicited by each probe type were balanced for length and frequency, we used the Kucera-Francis database of written frequency counts to calculate the frequency of the most popular name for each trial. The mean frequency was 71 (SD = 153) for cross-modal names and 100 (SD = 229) for speech-only names. Differences in the variance within each name type appear to derive from the occurrence of a few very high frequency words in response to speech-only pictures. Differences in mean written frequencies were not statistically reliable, t(335) = -1.4, n.s. The mean word length for cross-modal names was 5.8 (SD = 2.3) letters, and 5.6 (SD = 2.3) for speech-only names, t(335) = 0.63, n.s.

A second normative study evaluated whether related and unrelated items could be reliably interpreted as such. 20 additional volunteers listened to the digitized sound file extracted from each video clip and subjectively rated the degree of relatedness between each of the speaker's utterances and the subsequent picture probe on a scale of 1 to 5, with 5 designating the highest level. In the case of related trials, the mean rating was 4.3 (SD = 0.7) for cross-modal items, and 4.2 (SD = 0.8) for speech-only ones. In the case of unrelated trials, both sets of picture probes received mean ratings of 1.6 (SD = 0.6). As expected, related probes were rated as related to the prior context and unrelated probes were not. Moreover, when preceded by speech alone, cross-modal and speech-only related items were rated as equally related, and cross-modal and speech-only unrelated items were equally unrelated.

To rule out the possibility that ERP effects might derive purely from differences in visual complexity between cross-modal and speech-only stimuli, we recorded ERPs as six healthy adults (who did not participate in any other portion of this study) viewed picture probes presented in the absence of discourse primes. Two lists were constructed, each containing 84 cross-modal and 84 speech-only pictures.

Pictures were followed by single written words (identical to those used in the main experiment). Participants were instructed to attend to all pictures, and to classify each word as either related or unrelated to the preceding picture by means of a button press. ERPs elicited by cross-modal and speech-only pictures exhibited no visible differences for the entire course of an epoch extending from stimulus onset to approximately 1 s after. Mean amplitudes of ERPs to each probe type were measured within the time windows used in the main experiment to assess the N300 (250-350 ms after stimulus) and N400 (350-550 ms after stimulus). Additionally, the mean amplitudes of ERPs elicited by probe words that participants had subjectively classified as related or unrelated were also measured between 300 and 500 ms post-stimulus to assess lexical N400 effects.

Picture and word ERP measurements were subjected to a repeated-measures ANOVA with the factors of Stimulus Type (Cross-modal, Speech-only) and Electrode Site (see below for further explanation) for pictures, and Relatedness (Related, Unrelated) and electrode site for words. In response to pictures, no main effects of stimulus type or interactions with electrode site were found in either the time window for the N300 (main effect: $F[1,5]=.03$, n.s.; Stimulus × Electrodes interaction: $F[28, 140] = 0.8$, n.s.) or the N400 (main effect: $F[1,5] = 0.02$, n.s.; Stimulus x Electrodes interaction: $F[28, 140] = 0.8$, n.s.). However, in response to probe words, unrelated words reliably elicited more negative ERPs than related ones (Relatedness main effect: $F[1,5] = 7.4$, $p < .05$). This outcome demonstrates sufficient power in our sample size to detect reliable differences in brain response. Although interpreting null results yielded by the picture probes should nevertheless be approached with caution, the absence of any effect of stimulus type on ERP amplitudes suggests that the visual properties of cross-modal and speech-only stimuli were well balanced.

In order to encourage consistent attention to probe pictures, participants monitored for infrequent filler trials (7 total) in which the probes were dotted with blue paint splotches (applied through Adobe Photoshop). Data from these filler trials were not analyzed. Additionally, to provide an index of attention to discourse primes, each discourse-picture pair was followed by a single written task word that had either occurred in the immediately preceding speech, or was altogether new. Participants categorized each word as old or new by means of a button press, and response latencies and accuracy were measured, in keeping with Wu & Coulson (2005).

Four randomized lists were constructed, each containing 42 cross-modal related items and 42 speech-only related items. Each list also contained 84 unrelated trials, wherein cross-modal and speech-only probes (42 each) were paired with unrelated video clips. No video or probe picture was repeated on any list, but across lists, each picture appeared once as a related stimulus, and once as an unrelated one. Equal numbers of new and old task words followed each type of related and unrelated trial.

### ***Procedure***

Each trial began with a fixation cross, presented in the center of a 17 in. color monitor for one second. Video clips were presented at a rate of 48 ms per frame and varied in total length (mean = 3752 ms, SD = 1211 ms). After a 300 ms pause, a picture probe appeared on the screen for 400 ms. Nine hundred ms after the offset of the probe, the written task word was presented for 1 s (see Figure 4.3). A short pause (~ 5 s) followed each trial as the next set of video frames was loaded for presentation.

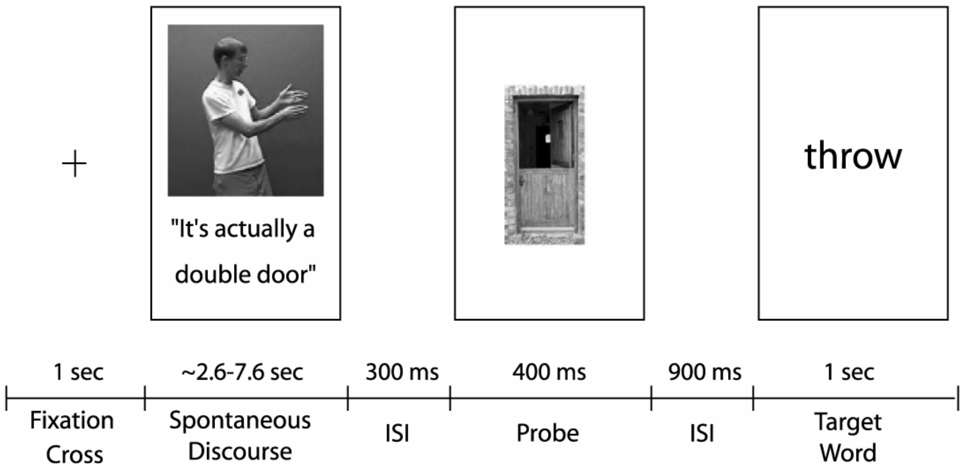


**Figure 4.3.** Procedure. Discourse primes were followed by picture probes, and then target words.

Participants were told that they would watch a short video of a man describing something, followed by a picture, and then a word. They were asked to press YES on a button box if they had heard the word uttered previously, or else to press NO. Response hand was counterbalanced across subjects. They were also instructed to monitor for infrequent blue splotches in picture probes, and were asked after each block if any had occurred.

### *EEG Recording*

The electroencephalogram (EEG) was recorded in a sound proof, electromagnetically shielded chamber. Tin electrodes were used 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) (see Figure 4.4). Electrodes were also placed on the right mastoid for off-line re-referencing, below the right eye for monitoring blinks, and a bipolar montage was placed at the outer canthi for monitoring horizontal eye movements. With the exception of the horizontal eye channels, all electrodes were referenced online to the left mastoid, and impedances maintained below five 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. Data were later re-referenced to the algebraic mean of the left and right mastoids.

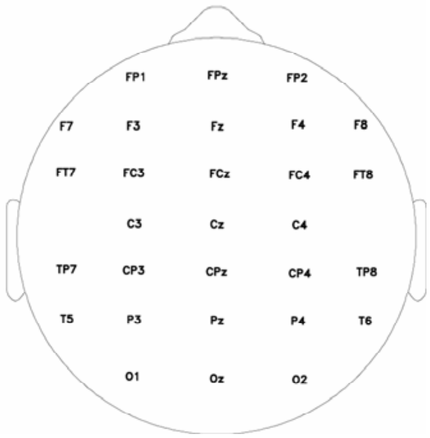


**Figure 4.4.** Schematic diagram of scalp electrode sites.

### *EEG Analysis*

Trials contaminated by artifacts such as blinks, eye-movements, blocking, and drift were rejected offline. Artifact-free trials were sorted and averaged, time-locked to the onset of picture probes. ERPs extended from 100 ms before stimulus onset to 920 ms after. On average, critical bins contained 35 trials (37 median). The mean artifact rejection rate was 14% percent (SD = 12%).

Relatedness effects were assessed by measuring the mean amplitude (that is, the average of digitized voltage measurements obtained within a sampling window, calculated relative to the pre-stimulus baseline) and peak latencies (i.e., the time point when the amplitude reaches its maximal value) of ERPs for each subject. Time windows for measurement were 250-350 ms (N300) and 350-550 ms (N400) after stimulus onset – based on measurement intervals utilized in other studies involving picture probes (Federmeier & Kutas, 2002; W. B. McPherson & P. J. Holcomb, 1999), as well as visual inspection of the waveforms. Measurements were subjected to a 2 × 2 repeated-measures ANOVA with the factors of Relatedness (probes were either related or unrelated to discourse primes), Stimulus Type (cross-modal probes depicted information made available in speech and gesture; speech-only probes depicted information expressed in speech alone).

To investigate the scalp distribution of ERP effects, an additional factor of Electrode Site (29 levels – corresponding to the 29 electrode channels) was included in the omnibus ANOVA. ERP effects qualified by an interaction with the electrode site factor were subject to three types off follow-up tests confined to specific groups of electrodes: Midline sites (with 7 levels along the anterior-posterior axis – namely, FPz, Fz, FCz, Cz, CPz, Pz, Oz), Medial sites (with 2 levels of Hemisphere – left and right – and 7 anterior-posterior levels – FP1/FP2, F3/F4, FC3/FC4, C3/C4, CP3/CP4 P3/P4, O1/O2), and from Lateral sites (with 2 levels of Hemisphere and 4 anterior-posterior levels – F7/F8, FT7/FT8, TP7/TP8, T5/T6). These follow-up analyses were designed to identify scalp regions where the effect was largest.

Additionally, we compared the topography of relatedness effects resulting from cross-modal and speech-only pictures by performing two point by point subtractions of ERPs elicited by related items from those elicited by unrelated items, yielding difference waves for each stimulus type. Repeated-measures ANOVAs were performed on both raw and normalized difference waves (McCarthy & Wood, 1985). (However, see Urbach & Kutas (2002) for counter arguments to the validity of vector scaling.) For all analyses, original degrees of freedom are reported; however, where appropriate, p-values reflect Geisser-Greenhouse correction (Geisser & Greenhouse, 1959).

## RESULTS

### *Behavior*

On average, participants accurately responded to 96% (SD = .03) of target words and 98% (SD = .02) of distractor words. A two-tailed *t* test revealed that this small difference was nevertheless reliable, t(15) = -2.5, p<.01, suggesting a slight bias on the part of participants toward the *no* response. The mean response time for classifying targets was 927 ms (SD = 322), and 1002 ms (SD = 316) for distractors. This difference did not approach conventional significance, however, t(30)=.67, n.s. –

perhaps due to insufficient power. Overall, the high accuracy rates and trend towards an advantage for targets suggests that participants consistently attended to video primes.

### *ERPs to Picture Probes*

Figure 4.5 shows the effect of relatedness in cross-modal and speech-only conditions. For both stimulus types, a negativity peaking around 130 ms (N1) can be observed, followed by two subsequent negativities, which peak around 295 ms (N300) and 430 ms (N400) respectively. ERPs elicited by unrelated items diverge from their related counterparts after 250 ms in the cross-modal condition, and after 350 ms in the speech-only condition. For both types of stimuli, ERPs remain more negative in response to unrelated items relative to related ones in the latter portion of the epoch (after 550 ms).

Does N300 sensitivity to relatedness differ for cross-modal pictures as compared to speech-only ones? Overall, unrelated stimuli elicited more negative ERPs than related ones between 250 and 350 ms post-stimulus (Relatedness main effect: $F[1,15] = 28$, $p < .0001$), and speech-only stimuli elicited more negative ERPs than cross-modal ones (Stimulus Type main effect: $F[1,15] = 10$, $p < .01$). However, these main effects were qualified by a Relatedness × Stimulus Type interaction, $F(1,15) = 4.5$, $p = .05$.

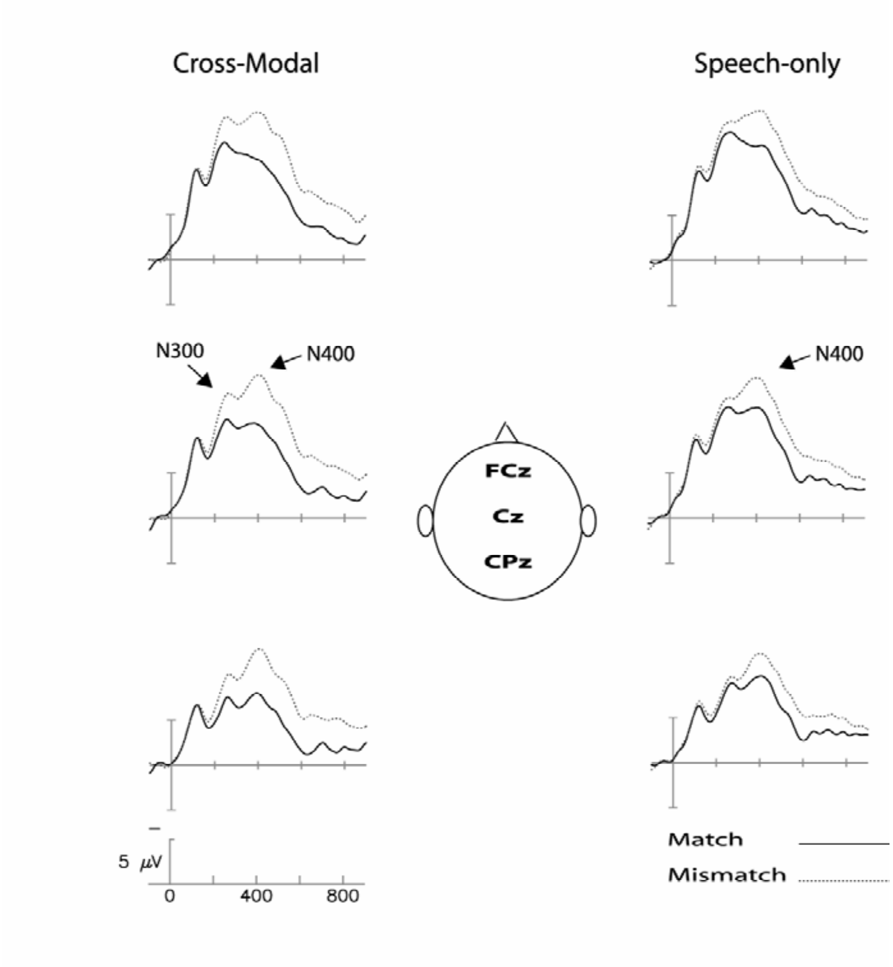


**Figure 4.5.** ERPs recorded over central midline sites time-locked to the onset of picture probes.

What drives the interaction between relatedness and stimulus type? Follow-up analyses of the effect of relatedness within each probe type revealed that cross-modal unrelated probes consistently elicited more negative ERPs than related ones (Relatedness main effect: F[1,15] = 51, p < .0001; Relatedness × Electrode Site: F[28,420] = 5.1, p < .005). For speech-only probes, by contrast, neither the main effect of relatedness nor the interaction with electrode site proved reliable, F’s < 2, n.s.

These outcomes indicate that the visuo-semantic processes indexed by the N300 reliably distinguished between related and unrelated items only in the case of cross-modal stimuli.

Where was the N300 effect in response to cross-modal stimuli largest? The interaction with electrode site obtained in the simple contrast between cross-modal related and unrelated items indicated that their effect on N300 amplitude was not uniform across the scalp. To characterize its distribution, follow-up analyses were conducted within midline, medial, and lateral electrode sites. In all three types of analyses, the N300 effect was most prominent over anterior electrode sites (Relatedness × Posteriority: Midline, F[6,90] = 3.6, p <.05; Medial, F[6,90] = 3.7, p = .07; Lateral, F[3, 45] = 10.5, p <.005) with a maximum over FCz and Cz . Further, the effect was larger over anterior right-hemisphere electrode sites than left-hemisphere ones (Relatedness × Hemisphere × Posteriority: Medial, F[6,90] = 2.9, p < .05) (See Figure 4.6 for scalp map of the N300 relatedness effect).

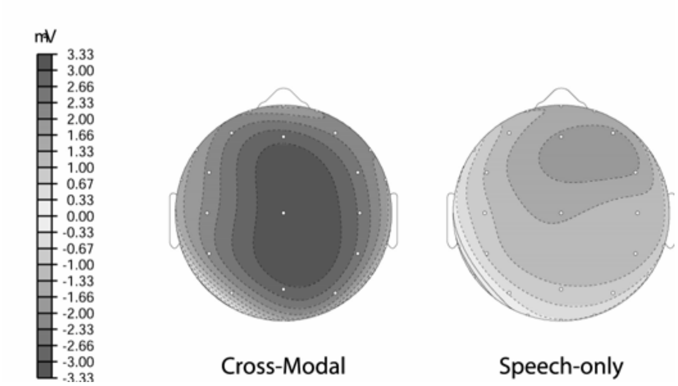


**Figure 4.6.** Scalp topography of N300 relatedness effect (Unrelated minus Related) at 300 ms post-stimulus onset. The N300 effect was significant only in the cross-modal condition.

Is the N400 also selectively sensitive to the semantic relatedness of cross-modal items? Between 350 and 550 ms, unrelated items consistently elicited more negative ERPs than related ones (Relatedness main effect: F[1,15] = 54.6, p <.0001). This main effect was qualified by an interaction between relatedness and stimulus type, F(1,15) = 4.7, p <.05. Follow-up analyses within cross-modal and speech-only probes revealed that unrelated items elicited more negative ERPs than related ones in response to both types of stimuli (Cross-modal: Relatedness main effect, F[1,15] = 57.7, p< .0001; Relatedness × Electrode Site, F[28, 420] = 10.3, p < .0001; Speech-only: Relatedness main effect, F(1,15)=23.4, p<.0005; Relatedness x Electrode Site, F(28, 420)=8.8, p<.0005). However, the relatedness effect was larger in the cross-modal condition (3 µV) as compared to the speech-only condition (1.9 µV ) (see Figure 4.7).

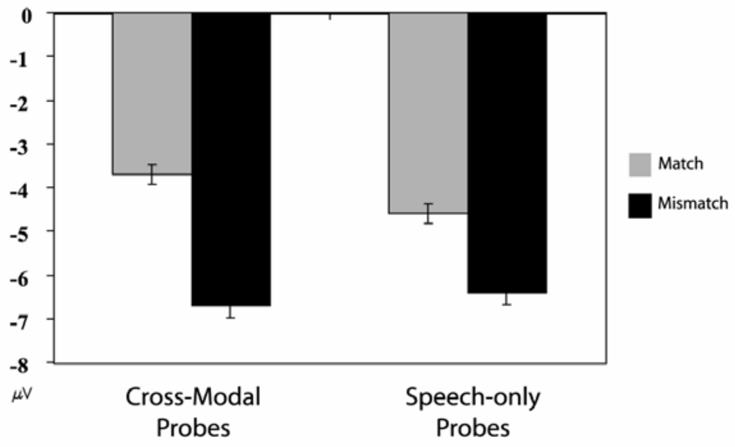


**Figure 4.7.** The mean amplitude of N400 across all electrode sites (in microvolts) elicited by cross-modal and speech-only picture probes.

What was the topography of cross-modal and speech-only N400 effects? Given relatedness by electrode site interactions obtained in responses to both stimulus types, follow-up analyses were performed to further assess the distribution of relatedness effects. For cross-modal items, this effect was largest over anterior electrodes, as is typical of N400 elicited by pictorial stimuli (Ganis et al., 1996; P. J. Holcomb & W. B. McPherson, 1994; West & Holcomb, 2002) (Relatedness × Posteriority: Midline, F[6, 90] = 9.3, $p < .001$); Medial, F[6, 90] = 8.7, $p < .005$; Lateral, F[3,45] = 9.6, $p <.01$) with a central midline maximum, as in the previously measured window. For speech-only items, the relatedness effect was also frontally focused (Relatedness × Posteriority: Midline, F(6, 90) = 7.7, $p< .005$; Medial, F[6, 90] = 6.4, $p < .01$; Lateral, F[3,45] = 4.3, $p = .05$). However, the effect was larger over the right than the left hemisphere (Medial: Relatedness × Hemisphere, F[1,15] = 6.3, $p < .05$; Relatedness × Hemisphere × Posteriority, F[6,90] = 5, $p < .01$; Lateral: Relatedness x Hemisphere, F[1,15] = 8.4, $p < .05$). (See Figure 4.8).

Do cross-modal and speech-only N400 effects reflect different distributions? Comparing difference waves of cross-modal and speech-only relatedness effects did not yield an interaction with electrode site during either the N300, F <1 n.s., or N400, F<1 n.s., time window for both raw and normalized data. This result indicates that the scalp distribution of relatedness effects elicited by each stimulus type did not reliably differ – despite subtle differences apparent in Figure 4.8.

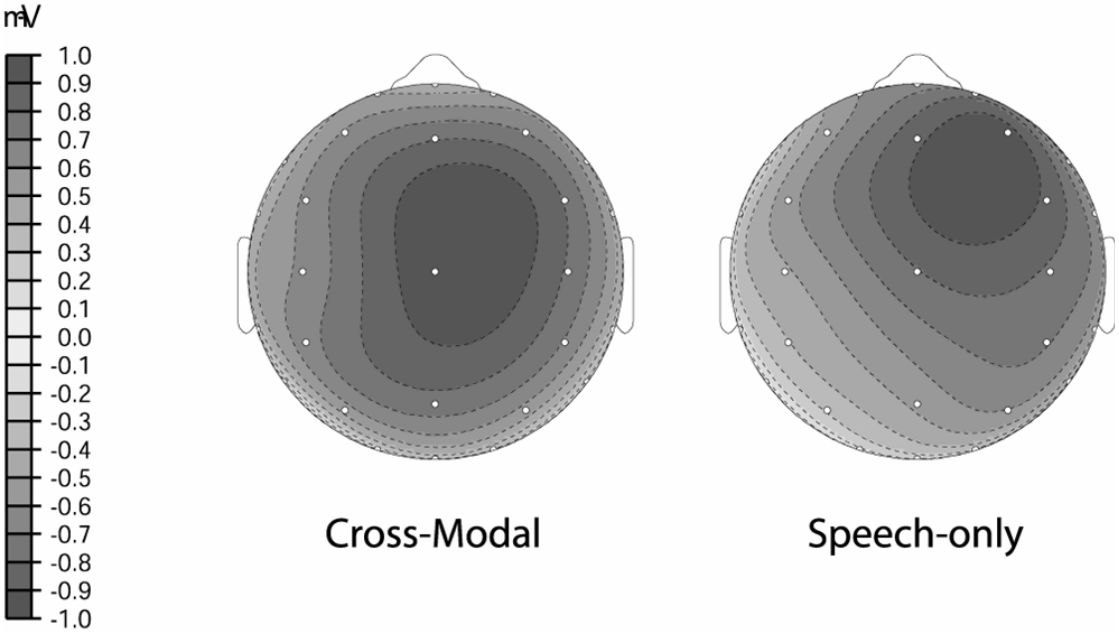


**Figure 4.8.** Spline maps depicting the distribution of the N400 effect (Unrelated minus Related) at 400 ms post-stimulus onset. For both cross-modal and speech-only conditions, values were normalized within a range of 1 and –1.

### *Post-Hoc Analyses*

Could differences between cross-modal and speech-only relatedness effects be driven purely by differences in the semantic fit between picture probes and the speech component of discourse primes? Although both cross-modal and speech-only probes were judged to be equally related to the speaker's utterances (see Methods), it is possible that speech segments were more predictive of cross-modal probes than of speech-only probes, resulting in the observed reduction of N300 and N400 in response to cross-modal related items relative to speech-only items. To rule out this possibility, we conducted a forced-choice normative study in which 18 additional volunteers listened to the sound-file extracted from each video-clip and indicated whether the speaker's utterance corresponded better to the cross-modal or speech-only related picture. On the basis of participants' responses, ERP trials were sorted into two categories. Cross-modally biased trials were those in which the cross-modal picture was preferred over its speech-only counterparts by more than 55% of respondents. Unbiased trials were those that garnered fewer than 55% of responses in favor of the cross-modal probe.

ERPs elicited by biased and unbiased pictures underwent 2 × 2 × 2 repeated-measures ANOVA with the factors of Bias, Relatedness, and Stimulus Type. Importantly, no main effect of Bias occurred within the time window of the N300, $F < 1$, n.s., or the N400, $F < 1$, n.s. Further, the factor of Bias did not interact with Relatedness or Stimulus Type within either time window, F's < 1, n.s. These findings suggest that the degree of semantic fit between speech segments and picture probes was not responsible for the larger relatedness effects observed in response to cross-modal as compared to speech-only items.

## Discussion

The brain response to cross-modal and speech-only probes differed in two ways. First, while cross-modal related items elicited less N300 than unrelated ones, no N300 effect was observed for speech-only probes. Secondly, although both related probe types elicited reduced N400 relative to unrelated controls, the N400 effect was larger to cross-modal probes. Below we discuss the implications of the cross-modal N300 effect for perceptual priming by gestures. Further, we discuss our finding of the larger cross-modal N400 effect relative to McNeill’s hypothesis that speech and gesture constitute an integrated system of thought.

### *N300 and Image Processing*

In the present study, discourse primes served to modulate the N300 response to cross-modal but not to speech-only picture probes. These findings suggest co-speech gestures affect image-specific semantic processes indexed by the N300. However, differing views have been advanced regarding the functional significance of this ERP component. McPherson and Holcomb (1999) propose that it reflects the activation of image-specific semantic relations between objects. Alternatively, given the finding that between-category violations (duck – collie) yield N300 effects, whereas within-category violations (poodle – collie) do not, the N300 has also been proposed to index a process whereby the structural properties of a percept are matched with a generic basic-level category representation before more identity-specific information becomes available (Hamm et al., 2002). Moreover, Schendan and Kutas have reported an early anterior negativity similar to the N300 that is larger in response to unidentified relative to identified objects (Schendan & Kutas, 2002), as well as in response to objects presented from unusual as compared to canonical views (Schendan & Kutas, 2003). These findings have led to the proposal that the amplitude of this anterior negativity is modulated by the size of the search space of possible object representations to which the percept could be matched, with larger amplitudes elicited by images that could correspond to a wide range of possible interpretations.

Ultimately, any inferences about the N300 must take into consideration the semantic richness of the stimulus conditions under which it is observed. When expectations regarding the content of an upcoming image are highly constrained by preceding context, N300 sensitivity to within-category violations has been reported (Federmeier & Kutas, 2001, 2002), in contrast to the findings of Hamm et al (2002). Similarly, in the present work, a differential N300 response was elicited by pictures that either did or did not match visuo-spatial cues made available through gestures. Cross-modal related pictures resulted in reduced N300 relative to unrelated ones, while the N300 elicited by speech-only related and unrelated items did not reliably differ. In other words, during this time interval the brain treats all speech-only probes as being completely unrelated to their preceding context, but differentiates between cross-modal related and unrelated items.

If the N300 reflects cognitive processing mediating object recognition, the present findings suggest that iconic gestures served to benefit the identification of cross-modal related stimuli. We propose that visuo-spatial cues provided by iconic gestures may have enabled listeners to formulate more precise conceptual representations of the items described in each utterance, thereby facilitating processes devoted to mapping percepts to stored knowledge and meaning for cross-modal related items, but not speech-only ones. This idea is consistent with current theories positing top-down facilitation of object recognition from low spatial frequencies in the image. It has been proposed that global shape information, such as orientation, size, and proportions, becomes available early during image processing, activating multiple possible high-level representations that constrain the interpretation of bottom-up input (Bar, 2003).

By analogy, visuo-spatial features of the gestures in this experiment may also have pre-activated a range of high-level representations that made related cross-modal items easier to identify than their speech-only counterparts. This hypothesis could be further investigated by augmenting the present experiment with a gesture-only condition, whereby related picture probes would agree with the speaker's gestures, but not his verbal utterances. If gestures exert top-down influences on probe image recognition, then larger N300 effects should occur in the gesture-only than in the speech-only condition.

### *N400 and Semantic Integration*

The enhanced N400 effect obtained in cross-modal trials relative to speech-only ones suggests that cross-modal related items were easier to interpret than their speech-only counterparts. Importantly, this processing difference can be attributed to the additional semantic cues supplied by gestures, as an off-line rating study revealed that both cross-modal and speech-only related items were rated as equally related to the speaker's utterances when his gestures were not shown. Additionally, when probe pictures were presented on their own, with neither speech nor gestures preceding them, both probe types elicited ERPs of similar amplitude during the N300 and N400 time windows. This finding discounts the possibility that ERP effects may have derived from differences in visual properties between cross-modal and speech-only stimuli.

The interaction between Stimulus Type and Relatedness in the present study was driven by a difference in the amplitude of the N400 elicited by the two types of related probes, whereas the unrelated probes elicited N400's of similar amplitude. This pattern suggests that the cross-modal N400 relatedness effect reflects facilitation of cross-modal related items rather than the detrimental impact of context on the processing of unrelated items. Inferences about these outcomes must be tempered, however, by the caveat that semantic integration processes indexed by the N400 may overlap temporally with object recognition processes indexed by the N300. Given this possibility, the greater magnitude of the cross-modal N400 effect relative to the speech-only one may be due at least in part to the differential magnitudes of the cross-modal and speech-only N300 effects. Again, an additional gesture-only condition in the present experimental paradigm would likely speak to this question. If the N300 and N400 reflect dissociable processes, images that are related only to the speaker's gestures should elicit less N300 activity than unrelated counterparts; however, no N400 relatedness would be expected.

Another issue deserving further exploration is the possibility that different relationships between gestures and speech used in the present experiment might affect listener comprehension in different ways. In some cases, gestures provided critical information denoting a certain kind of item within a class (e.g., a Dutch instead of a French door; a cupboard instead of a wall shelf; a stove knob instead of a door knob). In other cases, they portrayed salient visuo-spatial features of objects (e.g., the location of a logo on a T-shirt, the shape of vase, the degree of openness of a car window). Finally, some gestures demonstrated the manner of action execution (e.g., mixing with a spoon rather than an electric mixer, writing by hand rather than typing on a keyboard, painting with vertical rather than horizontal brush strokes).

It is possible that in response to cases where gestures provide substantive information beyond what is available through speech, listeners may formulate mental representations that are both visually and semantically more consistent with cross-modal probes relative to speech only ones. Consider an example in which the speaker demonstrates the shape of a tall, vertical cupboard while saying, "…and opposite that just kinda before the wall is another shelf." Here, any type of wall-mounted shelf would be congruent with the speaker's utterance alone, whereas a tall, upright cabinet or cupboard with shelves would be congruent with both his gesture and his speech.

On the other hand, in cases where the gesture simply elaborates information expressed through speech, it is possible that listeners activate representations that are visually more consistent with the cross-modal probe, but not semantically so. For example, in one trial the speaker says, "…a Nokia cell phone set at an angle," while indicating its orientation in gesture. In this instance, both the cross-modal and speech-only probes depicted the same cell phone, but at different orientations.

If the amount of additional information provided by gestures relative to speech affects speech-gesture integration, we might expect the two distinct types of discourse primes described above to differentially modulate ERP responses to picture probes. Because gestures that offer a great deal of additional information may result in more specific semantic activations (e.g. a tall cupboard with shelves instead of any type of shelves, or Dutch door instead of a French one), we might expect that when occurring in discourse primes, they may yield a larger N400 effect in the comparison between cross-modal and speech-only probes than in cases where the gesture served mainly to elaborate information expressed through speech. On the other hand, because both of these types of discourse primes allow the listener to formulate a more specific visual representation of the object or event being described than would be possible based on speech alone, we might expect comparably sized N300 effects for cross-modal versus speech-only probes in response to both prime types.

The present study confirms and extends existing experimental investigations of co-speech gesture integration. As noted above, researchers have previously reported that when speech and gesture convey different information, listeners are sensitive to both (Cassell et al., 1999; Goldin-Meadow et al., 1992). Further, listeners were shown to subsequently express gesturally conveyed meanings in speech, and vice versa, suggesting that activations from both modalities engage a common underlying substrate. The current study corroborates this view by demonstrating that semantic activations induced by speech and iconic gestures jointly contribute to emerging conceptual representations constructed during discourse comprehension. The fact that participants' on-going brain response to visual stimuli was modulated to a greater degree by cross-modal than speech-only stimuli suggests that listeners made use of semantic relations expressed through gesture, even though this information was never made overt in speech. This finding supports McNeill's proposal that during comprehension, listeners integrate both linguistically and gesturally encoded meanings. Additionally, an important theoretical consequence of this work is the idea that gestures enable listeners to construct perceptually specific conceptual representations of the speaker's intended message.

This proposal parallels sentence processing research that demonstrates that individuals make use of incoming linguistic input in order to formulate precise expectations about upcoming words. It has been shown, for example, that in sentences that strongly favor a particular kind of lexical completion, definite and indefinite articles that agree grammatically (Wicha, Bates, Moreno, & Kutas, 2003) or phonologically (DeLong, Urbach, & Kutas, 2005) with the anticipated completion elicit reduced N400 in comparison with those that do not. These results have been construed as evidence for pre-activation of specific word representations before their actual presentation.

By analogy, we propose that the gestures in the current study may pre-activate representations of visuo-spatial features including orientation, location, shape, and size, as well as motoric features associated with specific patterns of action execution. It is possible that mechanisms such as conceptual integration (Fauconnier & Turner, 1998, 2002) or "mesh" (Glenberg, 1997; Glenberg & Robertson, 1999) mediate a process whereby perceptual or relational similarities between the gesture and the entity being described make available the relevant visuo-spatial representations.

## Conclusion

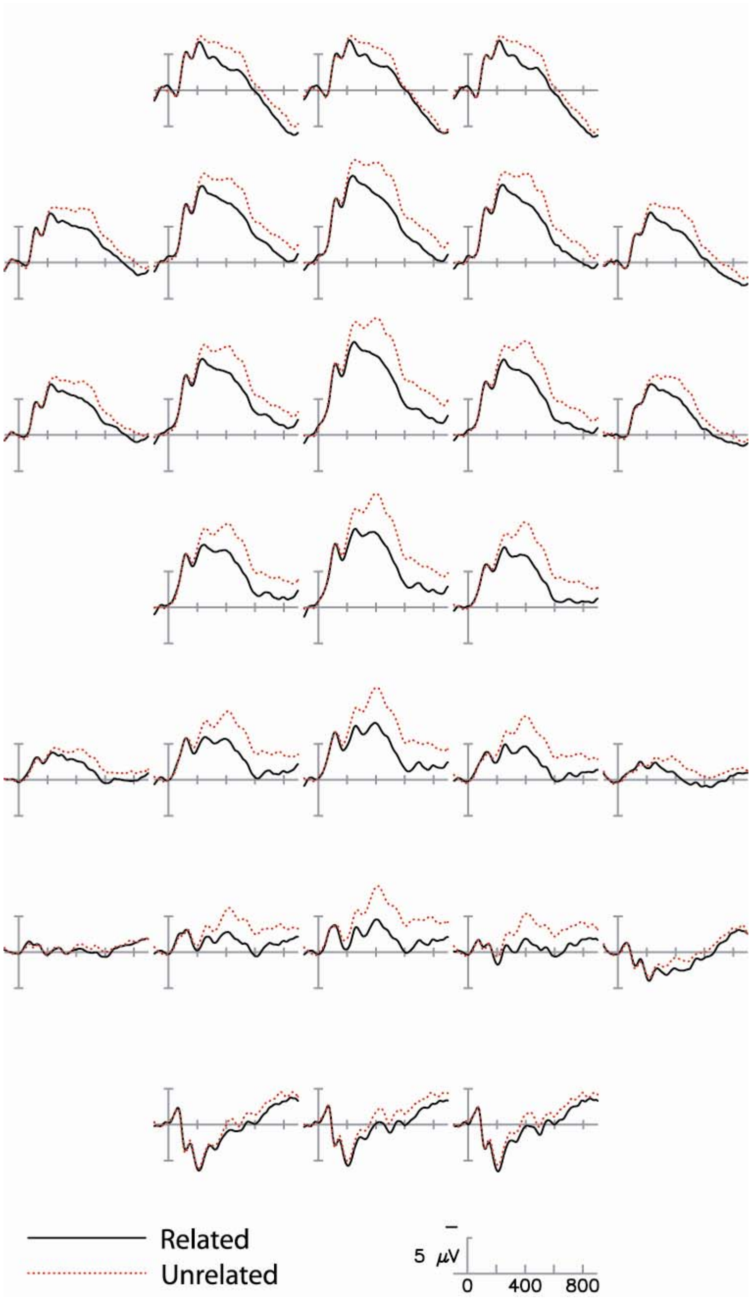
This study used ERPs to measure semantic activations prompted by co-speech gestures. We found that segments of spontaneously produced discourse involving speech and gesture differentially primed picture probes that agreed with information conveyed either through both channels or through speech alone. Cross-modal probes elicited a larger N400 relatedness effect than did speech-only probes. Additionally, cross-modal probes also elicited an N300 relatedness effect, whereas speech-only probes did not. These findings support the proposal advanced by McNeill (1992) that listeners combine information from speech and gestures to arrive at an enhanced understanding of their interlocutor's meaning. They further suggest that iconic gestures activate image-specific information about the concepts that they denote.

## Acknowledgement

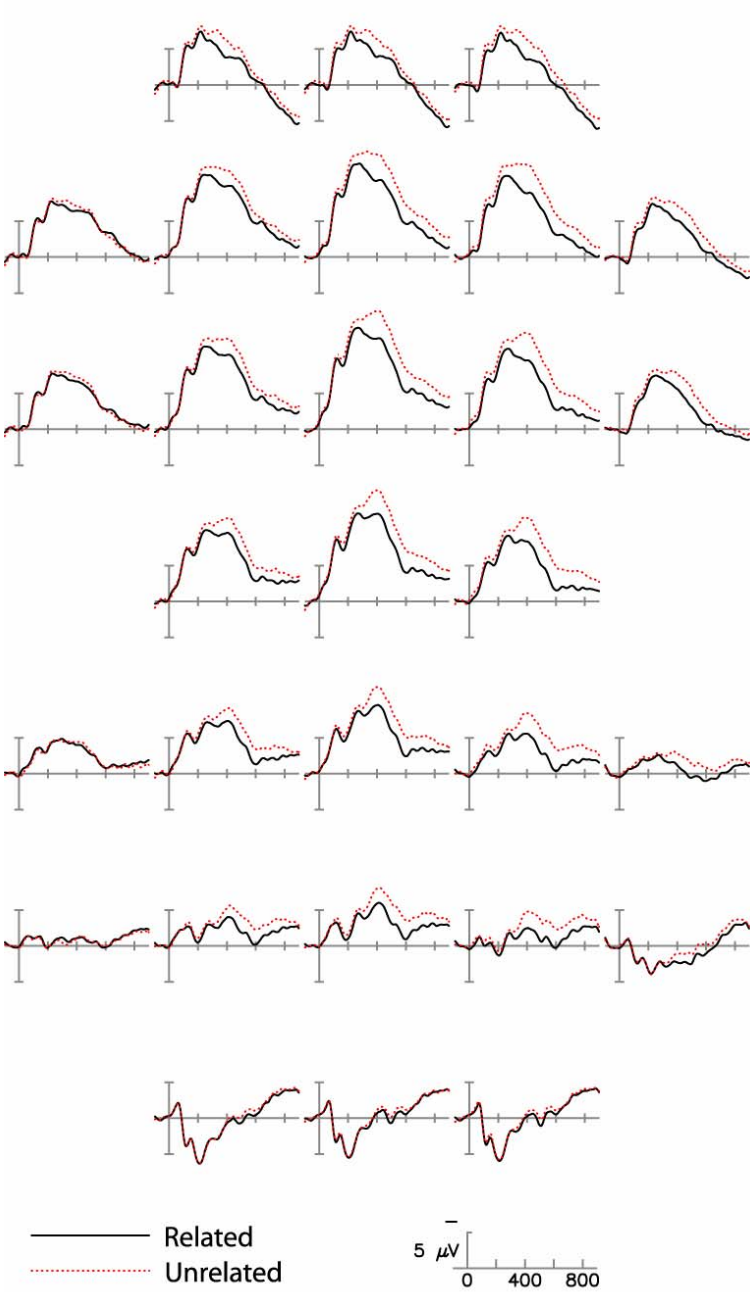
Y. W. was supported by Institute for Neural Computation NIH Training Grant # MH 2002-07 and Center for Research on Language NSF Training Grant DC 000-16.

Chapter 4, in full, has been accepted for publication of the material as it appears in *Brain and Language*, 2006, Wu, Ying Choon; Coulson, Seana, Elsevier, 2006. The dissertation author was the primary investigator and author of this paper.

# Appendix



Cross-modal N300 and N400 effects



Speech-only N300 and N400 effects

## Works Cited

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.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*(4), 600-609.

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. (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*(2), 179-192.

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), 1-33.

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

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilitstic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience, 8*(8), 117-1121.

Emmorey, K., & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture, 1*(1), 35-50.

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. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(202-224).

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: Eletrophysiological 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*(2), 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), 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.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge thorugh gestures: Using children's hands to read their minds. *Cognition and Instruction, 9*(3), 201-219.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology, 10*, 57-67.

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.

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

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.

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language, 28*, 325-349.

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., & Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*(1), 85-93.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 243-260.

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.

Kutas, M., Federmeier, K., Coulson, S., King, J. W., & Muente, T. F. (2000). Language. In G. G. Berntson (Ed.), *Handbook of Psychophysiology* (2nd ed., pp. 576-601). Cambridge, UK: Cambridge University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

McCarthy, G., & Wood, C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62*, 203-308.

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

McNeill, D. (1998). Speech and gesture integration. In J. Iverson & S. Goldin-Meadow (Eds.), *The nature and functions of gesture in children's communication. New directions for child development.* (Vol. 79, pp. 11-27). San Francisco: Jossey-Bass.

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

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

Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language, 19*(3), 559-580.

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating american sign language: Effects of sensory experience and age of acquisition. *Brain and Language, 57*(3), 285-308.

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.

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

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*, 188-194.

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.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience, 15*(1), 111-135.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85-89.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153-156.

Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology, 42*(1), 144-168.

Thompson, L. A., & Massaro, D. W. (1994). Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology, 57*, 327-354.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology, 39*, 791-808.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teacher's gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology 29*(187-204).

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*(3), 363-375.

Wicha, N. Y. Y., Bates, E., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165-168.

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

Wu, Y. C., & Coulson, S. (2005). Is that a meaningful gesture: Electrophysiological indices of gesture comprehension. *Psychophysiology, 42*(6), 654-667.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171.

# CHAPTER 5: GESTURE COMPREHENSION AND OBJECT RECOGNITION

Iconic gestures are spontaneous body movements that enhance on-going speech by depicting properties of objects and events. Research has demonstrated how these gestures have been used to communicate a wide range of semantic content, from basic visuo-spatial features, such as the direction of a block's rotation (Emmorey & Casey, 2001), to the operation of a forklift at a loading dock (Murphy, 2004), or procedures for manipulating specialized tools (LeBaron & Streeck, 2000). Yet, in spite of their commonness and semiotic richness, little is known about the underlying mechanisms mediating the comprehension of gestures such as these.

A number of theoretical views have either implicitly or explicitly pointed to similarities between the processes engaged during the comprehension of iconic gestures and those mediating the comprehension of pictures or bona fide objects. It has been argued, for example, that iconic 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, which suggests that access to meaning results from feature extraction and integration and from recognition of an invariant structure that does not depend on the viewer's perspective or on other peculiarities of executions (p. 90)."

The present study will explore the comprehension of iconic co-speech gestures and objects by comparing ERPs elicited during the semantic processing of these two stimulus types. The hypothesis that understanding gestures and objects involves at least partially overlapping cognitive systems gives rise to the prediction that specific ERP components implicated in object recognition should also be observed in response to gestures as well. Further, by comparing the distribution of ERP responses elicited by objects and gestures, it will be possible to draw inferences about the degree of overlap in the neural systems engaged by each of these representational media.

Previous electrophysiological research has uncovered two correlates of image-based semantic analysis. Specifically, in pictorial priming paradigms, contextually incongruent pictures have been shown to elicit an anterior negativity peaking around 300 ms after stimulus onset (N300), as well as a more broadly distributed negativity peaking approximately 400 ms post-stimulus (N400) (Barrett & Rugg, 1990; P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. Holcomb, 1999).

The N400 response to incongruous images has been related to the "classic" N400 elicited by 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 (Coulson & Federmeier, 2003; Wu & Coulson, in press). Expected sentence-final endings typically elicit little or no N400, for example, whereas unexpected completions elicit an N400 component with a large amplitude (Kutas &

Hillyard, 1980a, 1984). Likewise, sentences completed with unexpected pictures also result in a large N400 response as compared to expected ones (Federmeier & Kutas, 2001, 2002; Ganis et al., 1996; Nigam, Hoffman, & Simons, 1992).

In addition to sentential congruity, both the word and picture N400 are sensitive to different degrees of relatedness, with larger amplitudes for the second item of an unrelated pair (Kutas & Hillyard, 1988; W. B. McPherson & P. J. Holcomb, 1999). Just as pseudo-words elicit larger N400s than unrelated words (Holcomb, 1988), unrecognizable images elicit larger N400s than recognizable ones (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 or picture within a story context (van Berkum et al., 1999; West & Holcomb, 2002). Given these similarities in time course and functional characterization, the word and picture N400 have been proposed to reflect the activity of similarly functioning neural systems mediating the comprehension of language and images, respectively. However, N400 effects elicited by pictures tend to be larger over the front of the head, whereas N400 effects produced by words tend to be larger at centro-parietal sites (Ganis et al., 1996). This topographical difference is consistent with overlapping, but non-identical neural systems 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, but not words, also suggests some processes specific to image comprehension. Like the N400, the N300 is modulated by contextual congruity (Hamm et al., 2002), but is not sensitive to gradations of relatedness. An important study demonstrating this functional difference between the N300 and N400 was conducted by McPherson and Holcomb (1999). They presented viewers with image pairs exhibiting graded degrees of associative relatedness (e.g., highly related: toothbrush and toothpaste; moderately related: ketchup and mustard; unrelated: lion and stoplight). The amplitude of the N300 was found to differentiate between related and unrelated items, but not between moderately and highly related ones. That is, the N300 amplitude was shown to reflect whether items are related, but not the strength of the relation, since moderately related and highly related items had equally enhanced amplitude relative to unrelated items. By contrast, the amplitude of the N400 was found to exhibit sensitivity to gradations in relatedness between an image and its preceding context, with highly related items eliciting little or no N400 activity, moderately related items eliciting intermediate-sized N400s, and unrelated items eliciting even more reduced N400s.

An additional difference between N300 and N400 relatedness effects reported by McPherson and Holcomb (1999) was their distribution: N300 effects tended to be largest over anterior electrode sites, whereas N400 effects were more broadly distributed. On the basis of these functional and topographical differences, a number of researchers have proposed that the picture N300 and N400 index different aspects of the semantic processing of images. On the one hand, the N300 has been proposed to reflect the process of matching perceptual input with image-based representations in long-term memory (Schendan & Kutas, 2002; West & Holcomb, 2002). On the other, 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, iconic gestures have only recently begun to receive attention from cognitive neuroscience. A previous study from our lab (described in Chapter 2) tested the hypothesis that understanding iconic gestures is mediated by some of the same semantic integration processes responsible for the N400 in response to pictures and words (Wu & Coulson, 2005). A corpus of spontaneous, iconic co-speech gestures was produced by videotaping an individual describing cartoon segments and then using the recording to create short, soundless video clips of the speaker's gestures. The speaker was told that the experimenters were creating stimuli for a memory experiment and was unaware of the intent to elicit spontaneous gestures. Trials were constructed by pairing the original cartoon segments with either congruous or incongruous gesture clips. EEG was recorded from twenty four healthy adults as they viewed paired cartoon-gesture sequences.

ERPs time-locked to the onset of gesture presentation revealed that both congruous and incongruous gestures elicited a negative-going deflection of the waveform peaking around 450 ms after stimulus onset (gesture N450), with enhanced negativity for incongruous 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 N300 effects in preceding experiments may be interpreted as evidence that understanding gestures does not involve the kinds of image-specific activations indexed by the N300. Alternatively, it is also possible that due to the visual complexity of dynamic gestures as compared to static representations of objects, the processes indexed by the N300 are delayed in onset during gesture comprehension, overlapping with those indexed by the N450. 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, static gesture freeze frames were extracted from dynamic gesture video clips. Although static representations of iconic 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” is 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.

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 the original study, probe images were either related or unrelated to their preceding prime, or were unidentifiable. McPherson and Holcomb (1999) reported that unrelated probes elicited larger N300 and N400 than related ones, and likewise, unidentifiable items resulted in larger N300 and N400 than unrelated ones. However, N300 relatedness effects were larger over anterior electrode sites, whereas N400 effects were larger over centro-parietal ones.

This pattern of outcomes offers a baseline for comparing neural systems mediating gesture and object comprehension. If semantic processes recruited by these two types of stimuli involve overlapping systems, we would expect the distribution of N300 and N400 effects time-locked to single gesture jpegs to be similar to those elicited by object photographs. On the other hand, if these two stimulus types yield congruency effects with distinct scalp distributions, different configurations of neural generators will be implicated.

## Methods

### *Participants*

Twenty six 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 .735 (with maximal right handedness indicated by a score of 1). The data of three participants were excluded due to excessive artifacts, (greater than 40% of trials in critical bins).

When measurements from the remaining 23 participants were analyzed together as a single, comprehensive data set, no difference in N300 response to congruent and incongruent gesture trials was detected (F's < 2, n.s.). Because performance on the behavioral comprehension task varied considerably, ranging from near chance to perfect accuracy rates, it is likely that some participants did not consistently attend to stimuli. For this reason, the 7 worst comprehenders were parceled into a separate data set from the remaining 16 good comprehenders. Mean accuracy rates in response to unrelated words on gesture trials were 94% for both participant types (SD = 4%). However, high comprehenders on average accurately classified 86% (SD = 8%) of related words, whereas low comprehenders responded correctly on only 68% (SD = 11%) of related trials. ERPs and behavioral data from high and low comprehenders were analyzed separately.

### *Materials*

The gesture experiment contained static and dynamic items presented in an interleaved fashion. Dynamic gesture stimuli were taken from the same stimulus corpus described in Chapter 2. To create static gestures, we isolated a single frame (in the form of a jpeg file) from the video sequence of each gesture. Typically, jpegs were extracted just at the onset of the meaningful phase of movement – that is, the stroke (A. 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 extracted from the gesture stream. Congruous trials were those in which cartoon clips were paired with the original gesture produced while describing them. Incongruous trials involved mismatches. Cartoon-gesture pairs were followed by a task word, which participants were asked to judge either as related or unrelated (see Figure 5.1). 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 congruous and incongruous static trials were reliably distinguished as such. Six volunteers received academic credit for rating the degree of congruency between static gestures and cartoons on a five point scale. On average, the congruency rating was 3.3 (SD = 1) for congruent trials and 1.9 (SD = 0.4) for incongruent trials. 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 congruous and 80 incongruous 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. No cartoon, gesture, or word was repeated on any list, but across lists, each gesture appeared once as a congruous stimulus and once as an incongruous 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).

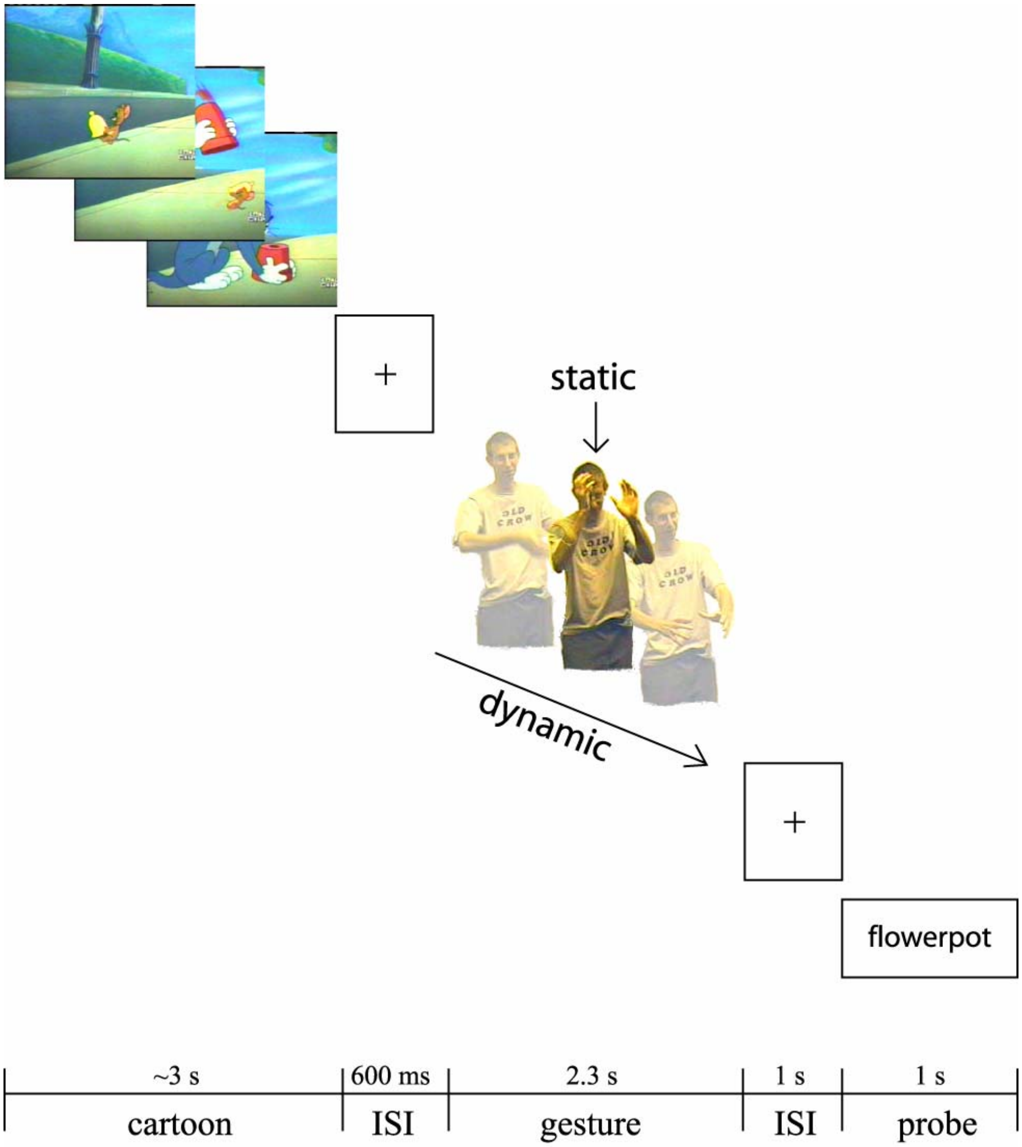


**Figure 5.1.** Trials included a short cartoon segment followed by either a dynamic or static gesture and then a probe word.

For the experiment involving photographs of common objects, 90 related (MOUSETRAP-MOUSE; BACON-EGGS) and 90 unrelated (BANANA PEEL-MOUSE; BASEBALL-EGGS) image pairs were constructed from digital photographs. Ninety additional trials were constructed by pairing photographs of common objects with unidentifiable targets (NOID's). 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). (For details on the selection and normalization of images, please consult McPherson and Holcomb (1999). 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 NOID 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 a NOID 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.

### ***Procedure***

Each volunteer 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 48 ms per frame with a 600 ms pause before the onset of the gesture (in order 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 (see Figure 5.1). A short pause (approximately 5 to 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.

Participants were told that they would watch a series of cartoon segments, each followed first by video clips 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 a practice block at the outset. Participants were presented with all eight blocks of gesture stimuli, followed by a short break, and then the final three blocks of object photos.

Picture trials also began with a 2 s fixation cross. The prime image was presented for one second, followed by a 250 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 500 ms, a fixation cross appeared for 500 ms, followed by the task word, which remained on the screen for one second. 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.

### *EEG Recording*

Recording parameters were identical to those described in Chapter 2.

### *EEG & Behavioral Data Analysis*

Participants' mean accuracy was assessed with repeated-measures ANOVA. For gesture trials, the factors of Gesture Type (Static, Dynamic), Gesture Congruency (Congruent, Incongruent), and Word Relatedness (Related, Unrelated) were used, yielding a 2×2×2 analysis. Picture trials were analyzed with the factors of Target Type (Related, Unrelated, or NOID) and Word Relatedness (Related, Unrelated).

Artifact-free ERP averages time locked to the onset of gestures were constructed from 100 ms before stimulus onset to 920 ms after. Trials accurately categorized by participants were sorted and averaged. Critical gesture bins contained 30 trials (±4) on average for high comprehenders, and 32 trials (±4) for low comprehenders. In the case of high comprehenders, the mean artifact rejection rate was 22% (12% SD) for dynamic gestures and 23% (10% SD) for static ones. In the case of low comprehenders, 17% (7% SD) of dynamic gesture trials and 18% (11%

SD) of static gesture trials were rejected. Critical picture bins contained on average 25 trials (±4) for high comprehenders, and 23 trials (±3) for low comprehenders. The mean artifact rejection rate for high comprehenders was 17% (13% SD) in response to identifiable picture probes, and 15% (13% SD) in response to NOID items. With regard to low comprehenders, 24% (10% SD) of identifiable picture probes and 20% (11% SD) of NOID probes were discarded. The higher rate of artifact contamination in the case of gestures relative to object photos may be due to increased ocular movements produced as dynamic gestures were viewed.

Gesture congruency and picture relatedness effects were both assessed by measuring the mean amplitude and peak latencies of ERPs time-locked to gesture onset from 300 to 400 ms (N300), 400 to 600 ms (Gesture N450), and 600 to 900 ms – in keeping with the intervals utilized in Chapter 2. Measurements were subjected to repeated-measures ANOVA, using within-subject factors of Gesture Type (Static, Dynamic), Congruency (Congruent, Incongruent), and Electrode Site (29 levels) for gesture trials, and within-subject factors of Target Type (Related, Unrelated, NOID) 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).

To compare the magnitude of ERP effects in response to static gestures and common objects for high comprehenders, a difference wave was constructed by performing a point by point subtraction of the averaged ERP waveform elicited by congruous trials from that elicited by incongruous 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, the N400, and extended processing effects, the mean amplitudes of static gesture and common object difference waves underwent repeated-measures ANOVA with the within-subject factors of Stimulus Type (Static Gesture, Common Object) and Electrode Site. Additionally, the N300 was measured a second time within the interval used by McPherson and Holcomb (1999) – namely 225 to 325 ms after onset. This second set of measurements was performed in order to obtain results that were more directly comparable with the outcomes reported by these researchers.

To investigate the scalp distribution of ERP effects deriving both from difference waves and raw averaged waveforms, we conducted 3 sorts of follow-up analyses: measurements of data collected from Midline sites (with 7 levels of anterior-posterior electrode location), measurements from Medial sites (with 2 levels of Hemisphere and 7 levels of anterior-posterior), and from Lateral sites (with 2 levels of Hemisphere and 4 levels of anterior-posterior).

## Results

### *Poor comprehenders*

As noted in the Methods section, participants were divided into two groups on the basis of their accuracy on the word classification task. We begin with analyses of data from the seven poor comprehenders, but will focus most of the discussion on the responses from high comprehenders.

### *Word Classification Accuracy*

Gesture Trials. In contrast to good comprehenders, who accurately classified 86% (SE = 8%) of related words on average, poor comprehenders classified only 68% (SE = 11%) of related words correctly. On average, poor comprehenders responded to unrelated words (94% SE = 1%) more accurately than related ones (68% SE = 3%) (Relatedness main effect: F[1,6] = 38.5, p < .005). Further, words following congruent cartoon-gesture pairs were classified more accurately (86%, SE = 2%) than those following incongruent ones (77%, SE 2%) (Congruency main effect: F[1, 6] = 31, p< .005). Main effects were qualified by a three way interaction between Gesture Congruency, Word Relatedness, and Stimulus Type, F(1, 6) = 7, p< .05. This interaction was driven by the fact that participants' mean accuracy rates in response to words following incongruent static and dynamic gesture trials were approximately 20% to 40% lower than in all other conditions.

Object Photo Trials. Analyses of the word classification task on object photo trials revealed a trend toward an interaction between Word Relatedness and Target Type, F(2, 12) = 3.2, p = .08. Related words were classified much more accurately following related picture targets (93%, SE = 3%) than unrelated (72%, SE 5%) or unidentifiable (70%, SE = 9%) targets.

### *ERPs*

Gestures. Neither the N300 nor N400 was consistently modulated by the congruency of gestures or their mode of presentation, F's < 2, n.s. A trend towards an interaction between Gesture Type and Electrode Site was observed between 400 and 600 ms post-onset, F(28, 168) = 3, p = .06, reflecting the greater negativity of ERPs elicited by static trials relative to dynamic ones over posterior electrode sites. Finally, testing between 600 and 900 ms post-onset revealed a reliable interaction between Gesture Congruency and Stimulus Type, F(1, 6) = 13, p< .05. However, testing separately within each gesture type did not yield any reliable congruency effects, F's < 2.75, n.s.

Object Photos. Between 300 and 400 ms post-stimulus, marginally significant effects of Target Type were observed (main effect: F[2, 12] = 3.5, p = .07; Target Type × Electrodes Interaction: F[56, 336] = 6.5, p = .05). A pre-planned contrast revealed that unrelated targets elicited N300s of greater magnitude than related ones, but this effect only approached conventional significance, F(28, 162) = 4, p = .09. Likewise, unidentifiable targets resulted in larger N300s than unrelated ones over anterior scalp electrode (Target Type × Electrodes Trend: F[28, 168] = 3, p < .08).

The amplitude of the N400 was consistently modulated by the relatedness of object photo targets (Target Type × Electrodes: F[56, 336] = 3, p < .05). Unrelated objects elicited greater N400s than related ones across the scalp, F(1, 6) = 8.4, p <.05), whereas unidentifiable items elicited more N400 activity than unrelated objects primarily over anterior recording sites, (Target Type × Electrodes, F[28, 168] = 3.7, p

<.05; Midline: Target Type × Posteriority, F[6, 36] = 6, p <.05; Medial: Target Type × Posteriority, F[6, 36] = 4.8, p < .05).

Finally, between 600 and 900 ms, sensitivity to the semantic properties of targets was still evident (Target Type trend: F[2, 12] = 3.9, p = .05). Unrelated trials continued to elicit more negative ERPs than related ones, F(1,6) = 8.4, p < .05; however, unidentifiable trials did not differ reliably from unrelated ones, F's < 1.5, n.s.

### ***High Comprehenders***

#### *Word Classification Accuracy*

<u>Gesture Trials.</u> On average, participants classified 87% of related words and 93% of unrelated words correctly. The condition of Gesture Type (static versus dynamic) yielded no main effect, F(1,15)=1.6, n.s., or interactions, F's < 1.5, n.s. By contrast, a main effect of Gesture Congruency indicated that words were classified more accurately when following congruent cartoon-gesture pairs relative to incongruent ones, F(1, 15) = 20, p < .0005. Additionally, a main effect of Word Relatedness indicated greater accuracy for unrelated words relative to related ones, F(1, 15) = 8.5, p < .05. These main effects were qualified by an interaction (Gesture Congruency × Word Relatedness, F[1, 15] = 17.5, p < .001). Further inspection of the data revealed that participants responded less accurately to related words following incongruous cartoon-gesture pairs than to any other word type (see Table 5.1).

**Table 5.1.** Percentage of accurately classified words following congruent and incongruent gestures.

| | Congruent Gesture | Incongruent Gesture |
|---|---|---|
| Related Word | 92% (8% SE.) | 81% (12% SE.) |
| Unrelated Word | 93% (7% SE.) | 93% (6% SE.) |

Object Photo Trials. Participants' mean classification accuracy was 83% for related words and 92% for unrelated ones. A main effect of Word Relatedness confirmed that this difference was reliable, F(1, 15) = 16, p < .005. Further, a main effect of Target Type, F(2, 30) = 20, p < .0001, indicated that words following related picture pairs were classified more accurately (94%) than words following unrelated (84%) or NOID trials (83%). However, the trend towards an interaction between these factors only approached conventional significance, F(2,30) = 3, p = .09.

*ERPs*

Figure 5.2 shows ERPs recorded over the midline 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 240 ms after stimulus onset, followed by a second negative-going deflection of the waveform peaking

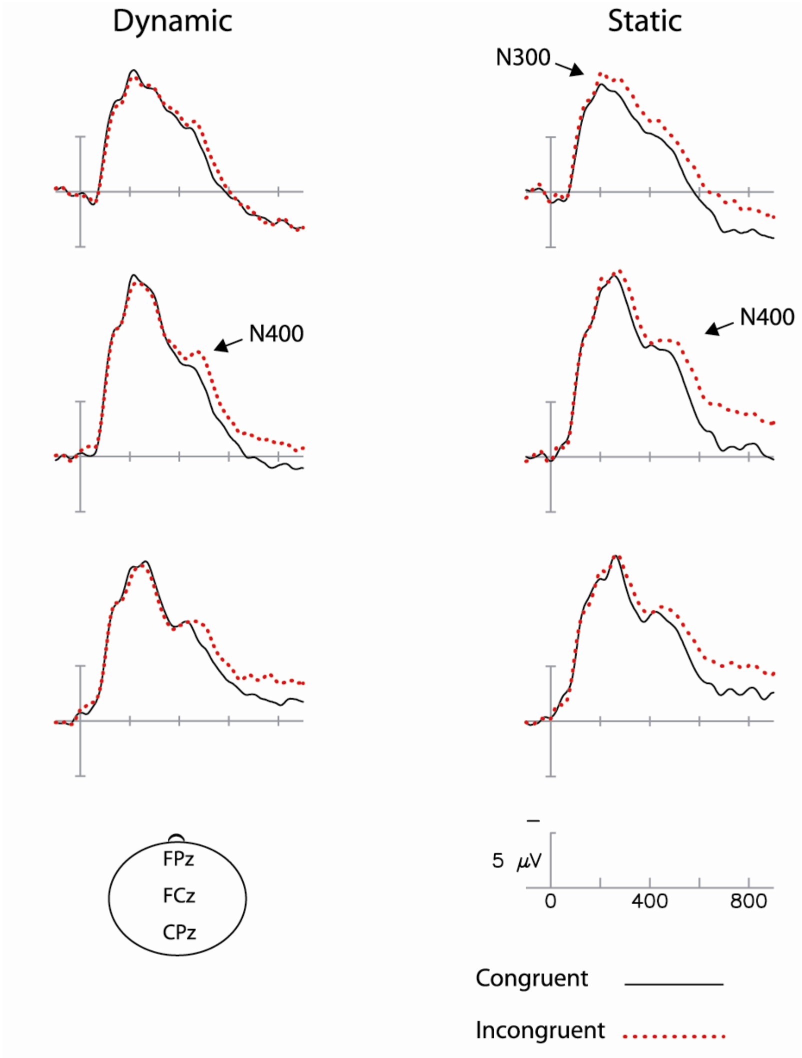


**Figure 5.2.** ERPs time-locked to the onset of static and dynamic gestures. Note that negative voltage is plotted up.

around 450 ms post-stimulus. In the case of dynamic gestures, effects of gesture congruency can be observed from around 400 ms to the end of the epoch (900 ms), with enhanced negativity for incongruous items relative to congruous ones. In the case of static gestures, congruency effects begin earlier, around 230 ms after stimulus, again with incongruous items continuing to elicit more negative ERPs until the end of the epoch.

Figure 5.3 depicts ERPs elicited by all three types of image pairs, recorded idline and medial electrode sites. The negative-going onset potential visible in all conditions peaks around 100 ms after stimulus onset, and is smaller than that observed in response to gestures. Effects of image relatedness are visible by 100 ms post-stimulus and continue to the end of the epoch, with unrelated image pairs eliciting more negative ERPs than related ones. NOID trials are differentiated from unrelated ones between approximately 200 and 600 ms after stimulus onset, with NOID items resulting in more negative ERPs than unrelated items.

N300 Time Window. For gesture trials, an omnibus ANOVA analyzing the mean amplitude of ERPs measured between 300 and 400 ms post-stimulus did not yield main effects of Gesture Type or Gesture Congruency, F's < 2, n.s. However, an interaction between these two factors was observed, F(1,15) = 5.7, p < .05. To determine the cause of this interaction, follow-up comparisons of congruency effects were conducted individually within each type of gesture. For static gestures, incongruous trials consistently resulted in more negative ERPs, F(1,15) = 6, p < .05. However, for dynamic trials, gesture congruency did not reliably modulate ERP amplitudes (F = .16, n.s.) during this epoch. This pattern of outcomes is consistent with Figure 5.2, which shows a distinct N300 effect elicited by static gestures, but no N300 differences between congruent and incongruent dynamic gestures.

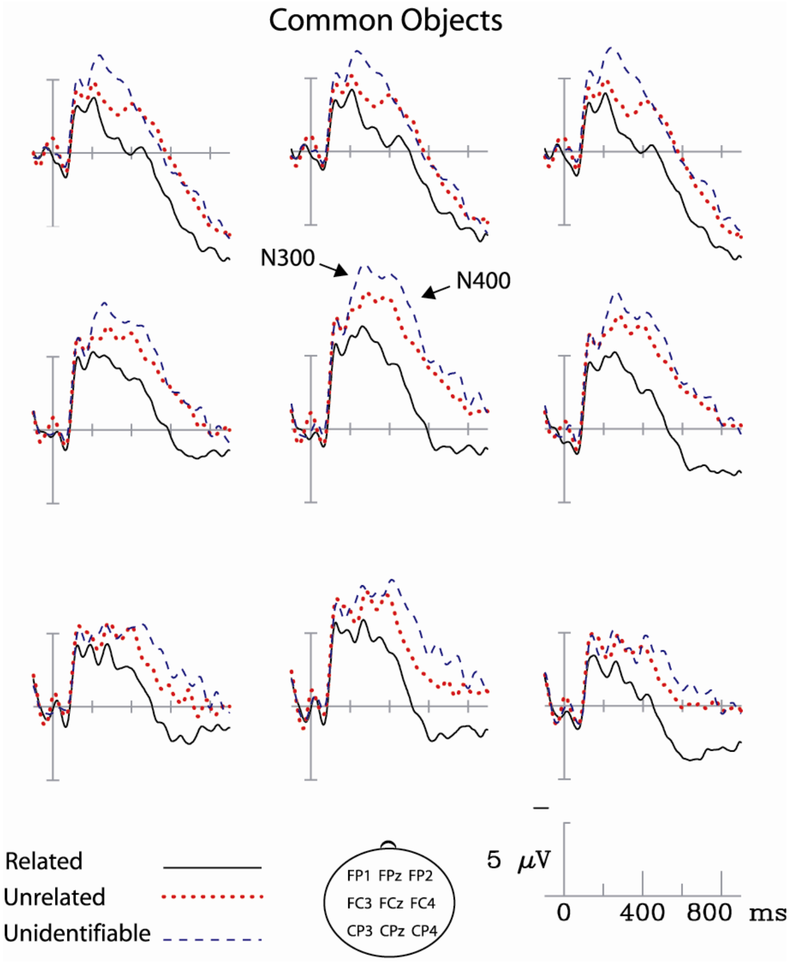


**Figure 5.3.** ERPs elicited by related, unrelated, and unidentifiable photographs of common objects.

To assess how image relatedness modulated the N300, the mean amplitudes of ERPs elicited by related, unrelated, and NOID trials were compared, measuring between 300 and 400 ms post-stimulus. A main effect of Target Type, F(2, 30) = 12, p < .0005, along with a Target Type × Electrode Site interaction, F(56, 840) = 11.5, p < .0001, 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 NOID pairs.

Contrasting related and unrelated image pairs demonstrated, as expected, that unrelated trials consistently resulted in larger N300s than related trials (see Table 5.2 for outcomes of this contrast and subsequent analyses investigating the scalp distribution of the relatedness effect). This effect was larger over anterior scalp electrodes and over right hemisphere sites. Contrasting unrelated and NOID trials revealed that unidentifiable images resulted in more negative N300s than unrelated images over anterior but not posterior electrode sites (see Table 5.2).

**Table 5.2.** Contrasts of mean amplitude of ERPs elicited by related, unrelated and NOID object photos between 300 and 400 ms post-stimulus.

| | Target Type | Target Type x Electrodes | Target Type x Posteriority | | | Target Type x Hemisphere | | Target Type x Hemisphere x Posteriority | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Midline | Medial | Lateral | Medial | Lateral | Medial | Lateral |
| Contrast | F(1,15) | F(28, 420) | F(6, 90) | F(6,90) | F(3,45) | F(6,90) | F(3,45) | F(6,90) | F(3,45) |
| Related vs Unrelated | 21.3** | 4.2* | 4.6* | 4.0* | n.s. | n.s. | 5.5 | 3.6 | n.s |
| Unrelated vs NOID | n.s. | 9.0** | 7.0* | 12.5* | 22.0** | n.s. | n.s. | n.s. | 6.3 |

* p < .05

** p < .0005

N400 Time Window. For the gesture experiment, analysis of measurements obtained between 400 and 600 ms post-stimulus revealed a main effect of Gesture Congruency – collapsed across both types of gesture, F(1,15)=12.5, p<.005. This effect occurred due to consistently larger gesture N450 responses to incongruent relative to congruent trials.

The main effect of Gesture Congruency was qualified by an interaction with Electrode Site, F(28,420) = 3, p <.05, indicating that effects were not uniform across the scalp. The distribution of the N450 congruency effect was investigated with follow-up analyses at midline, medial and lateral electrode arrays. At the midline, a trend toward a Congruency x Posteriority interaction( F[6,90] = 3, p = .07) was driven by the increased congruency effect over frontal electrode sites relative to posterior ones for both static and dynamic gestures. For both gesture types, the congruency effect was also larger over right than left hemisphere sites than over left hemisphere sites (Gesture Congruency × Hemisphere: Medial, F[1,15] = 3.5, p = .08; Lateral, F[1,15] = 4.9, p < .05).

In addition to effects of Gesture Congruency, a main effect of Gesture Type also proved reliable, F[1,15] = 8, p < .05. This outcome resulted from the more negative mean amplitude of ERPs elicited by static gestures relative to dynamic ones, irrespective of congruency. An interaction between Gesture Type and Electrode Site was also obtained, F(28, 420) = 3.8, p < .05. Follow-up analyses revealed that the effect of Gesture Type was larger over central midline and medial electrode sites (Gesture Type × Posteriority: Midline, F(6, 90) = 7, p < .005; Medial, F(6, 90) = 5, p < .05).

For the object photo experiment, an analysis of measurements obtained within the same time window revealed a main effect of Target Type, F(1,15) = 20.6, p < .0005, and an interaction with Electrode Site, F(56,840) = 7, $p < .0001$. A simple contrast between related and unrelated items indicated that unrelated items elicited larger N400s than related items (see Table 5.3). As with the N300, this effect was frontally focused, and larger over the right than the left hemisphere. The contrast between unrelated and unidentifiable images established that NOID items resulted in larger N400s than did unrelated images – again chiefly over anterior electrodes.

**Table 5.3** Contrasts of mean amplitude of ERPs elicited by related, unrelated and NOID object photos between 400 and 600 ms post-stimulus.

| | Target Type | Target Type x Electrodes | Target Type x Posteriority | | | Target Type x Hemisphere | | Target Type x Hemisphere x Posteriority | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Midline | Medial | Lateral | Medial | Lateral | Medial | Lateral |
| Contrast | F(1,15) | F(28, 420) | F(6, 90) | F(6,90) | F(3,45) | F(6,90) | F(3,45) | F(6,90) | F(3,45) |
| Related vs Unrelated | 66.6** | 6.0* | 4.0* | 6.2* | 3.3 | 5.0* | 12.0* | 2.6 | n.s. |
| Unrelated vs NOID | n.s. | 3.5* | 3.8* | 4.2* | n.s. | n.s. | n.s. | n.s. | 17.2** |

* p < .05

** p < .0005

<u>600 to 900 ms post-stimulus.</u> For the gesture experiment, a final analysis assessing on-going ERP effects beyond the N400 time window revealed a main effect of Gesture Congruency, F(1,15) = 13.6, p < .005, but not Gesture Type, F = 2.8, n.s. These outcomes were qualified by an interaction between Gesture Congruency and Gesture Type, F(1,15) = 5.2, p < .05. Follow-up ANOVAs conducted separately within static and dynamic gestures revealed that dynamic gestures did not yield reliable congruency effects from 600 to 900 ms post-onset, F = 1.7, n.s. By contrast, in the case of static gestures, incongruent trials elicited reliably more negative ERPs than their congruent counterparts up to the end of the epoch (Gesture Congruency main effect: F[1,15] = 26, p < .0005; Gesture Congruency × Electrode Site: F[28,420] = 4.25, p < .01).

Because the simple contrast between congruent and incongruent static gestures (600-900 ms) was qualified by an interaction with electrode site, post hoc follow-up tests were conducted to investigate the scalp distribution of the static gesture congruency effect. Interactions between Gesture Congruency and Posteriority indicated that the effect was most prominent over anterior electrode sites (Midline: F[6,90] = 5.3, p < .05; Medial: F[1,15] = 3, p = .09; Lateral: F[3,45] = 4.4, p < .05). Additionally, a trend was found suggesting that the static gesture congruency effect was larger over right hemisphere electrodes (Gesture Congruency × Hemisphere: Lateral, F[1,15] = 3.5, p = .08).

For the object photo experiment, final analysis on ERPs measured between 600 and 900 ms post-stimulus revealed on-going sensitivity to image relatedness (Target Type main effect, F[1, 15] = 19.1, p < .0005; Target Type × Electrodes

Interaction, F[56, 840] = 3.7, p < .005). Unrelated trials continued to elicit more negative ERPs than related ones. This effect was largest over anterior right hemisphere electrode sites (see Table 5.4). Unidentifiable trials, however, were no longer reliably distinguished from unrelated ones, F's < 1, n.s.

**Table 5.4.** Contrasts of mean amplitude of ERPs elicited by related, unrelated and NOID object photos between 600 and 900 ms post-stimulus.

|  | Target Type | Target Type x Electrodes | Target Type x Posteriority |  |  | Target Type x Hemisphere |  | Target Type x Hemisphere x Posteriority |  |
|---|---|---|---|---|---|---|---|---|---|
|  |  |  | Midline | Medial | Lateral | Medial | Lateral | Medial | Lateral |
| Contrast | F(1,15) | F(28, 420) | F(6, 90) | F(6,90) | F(3,45) | F(6,90) | F(3,45) | F(6,90) | F(3,45) |
| Related vs Unrelated | 68.0** | 3.7* | n.s. | 4.9* | n.s. | 5.9* | 10.0* | 3.0 | n.s. |
| Unrelated vs NOID | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | n.s. | 14.5** |

* p < .05

** p < .0005

*Comparison between static gesture and object photo ERP effects.* Measuring within the time window used by McPherson and Holcomb (1999) to assess the N300 (225 to 325 post-stimulus), the magnitudes of the static gesture congruency effect and the object photo relatedness effect did not consistently differ (F's < 1.5, n.s.). However, within the time window used to assess the N300 in the present study (300 to 400 ms post-stimulus), a trend toward a main effect of Stimulus Type was observed, F(1, 15) = 3.9, p = .07. Post-hoc follow-up analyses confined to medial and lateral recording sites revealed that this effect was larger over anterior right hemisphere electrodes relative to left hemisphere counter parts (Stimulus Type × Hemisphere × Posteriority: Medial, F[6,90] = 3.4, p < .05). Subsequently, between 400 and 600 ms post-stimulus, a reliable main effect of Stimulus Type was obtained, F(1, 15) = 15.4, p< .005); however, post-hoc follow-up tests did not yield any reliable interactions with Hemisphere or Posteriority.

During both the measurement interval for the N300 and the N400, 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 N400 effect, and possibly a larger, more right lateralized N300 effect. Finally, from 600 to 900 ms post-stimulus, no differences were observed between the size of congruency effects elicited by each stimulus type, F < 2.5, n.s.

## Discussion

To investigate the online processing of gestures, we recorded ERPs time-locked to the onset of spontaneously produced, iconic gestures preceded by congruent and incongruent contexts. In addition to congruency, 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. Incongruous static gestures elicited an enhanced, frontally focused negativity peaking between 300 and 400 ms post-stimulus (N300). Between 400 and 600 ms post-stimulus, all gestures elicited a negative deflection of the ERP waveform (gesture N450) with incongruous items yielding enhanced negativity relative to congruous ones. Both types of gestures also elicited a late congruency effect (600-900 ms), which was much larger, more robust, and more broadly distributed for static items.

In addition to gestures, ERPs were recorded in response to related and unrelated pairs of photographs depicting common objects. 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. In the final portion of the epoch (600-900 ms), unrelated items continued to elicit more negative ERPs than related ones, but were not reliably distinguished from unidentifiable ones.

### *Gesture ERP effects*

The gesture N450 effect described here corroborates outcomes reported in previous studies using this paradigm (Wu & Coulson, 2004). It 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; P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. Holcomb, 1999; West & Holcomb, 2002). Further, the time course of the N450 effect observed in the present study was in keeping with that of N400 activity reported in experiments involving visually complex scenes (West & Holcomb, 2002), or videographic stimuli (Sitnikova et al., 2003).

The new finding produced by the present study is the modulation of the N300 by gesture congruency. The N300 is a frontally distributed negativity 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 (W. B. McPherson & P. J. Holcomb, 1999), photographs of conventionally meaningful hand shapes (such as “thumbs up”) (Gunter & Bach, 2004), and complex scenes (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 (P. J. Holcomb & W. B. McPherson, 1994; W. B. McPherson & P. J. Holcomb, 1999).

On the basis of this functional characterization, the N300 has been linked to a similar 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 authors conclude that this ERP component indexes object model selection, whereby the perceived image is compared with possible structural representations stored in long term memory. Accordingly, unidentified pictures elicit a large negativity due to the increased search space engendered by the greater range of potential matches. By contrast, the amplitude of this negativity is reduced when a percept may be readily reconciled with a corresponding object model, as in the case of identified items.

In the present study, the finding of reduced N300 in response to contextually congruous static gestures suggests that corresponding stored structural representations were easier to access in the case of congruous relative to incongruous items. This idea is particularly exciting because it suggests that the cognitive system mediating object recognition is sensitive to semiotic properties of iconic gestures. 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 the semiotic properties of congruent gestures did not affect object recognition processes, then the brain would be expected to respond to both congruent and incongruent stills as essentially equally unrelated images of a man, and no N300 effect would be predicted.

Importantly, as discussed in Chapter 2, 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. His actions cause the candle to topple, much in the manner of a tree being felled, onto Jerry's head. The subsequent gesture still is shown in Figure 5.4, along with a single freeze from the cartoon. 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 5.4 will confirm, there are few similarities between the cartoon

**Figure 5.4.** A freeze frame extracted from a cartoon segment used in the experiment, and the subsequent static gesture.

and the gesture from a purely perceptual perspective. For example, 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, whereas the speaker's arm is covered by a the sleeve of a plaid shirt.

If the N300 indexes image specific processes critical to gesture comprehension, the question arises why congruency effects within this time window are observed only in response to static and not dynamic gestures. Or from a different perspective, one might ask whether the static gesture N300 generalizes to natural gesture processing, or is simply an artifact of the use of freeze frames. These concerns can be countered by the possibility that in response to moving, visually complex stimuli, processes indexed by the N300 might become activated slightly later than they would in response to static objects, overlapping with processes 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. Rather, it may be the case that the activation and integration of image-based representations are concurrent cascading processes.

Support for this idea can be found in Figure 5.5, which compares isovoltage and current source density maps (CSD) of the N300 and N450 effects for both static and dynamic stimuli. At the approximate time points within the N300 and N450 windows when the difference between congruent and incongruent stimuli was maximal, interpolated values of the congruency effect were plotted (in microvolts) across the scalp. Current source density maps were computed from isovoltage maps, revealing an estimation of sources and sinks of radial current. For each map, normalization was implemented by dividing each data point by the maximum absolute value for that plot.

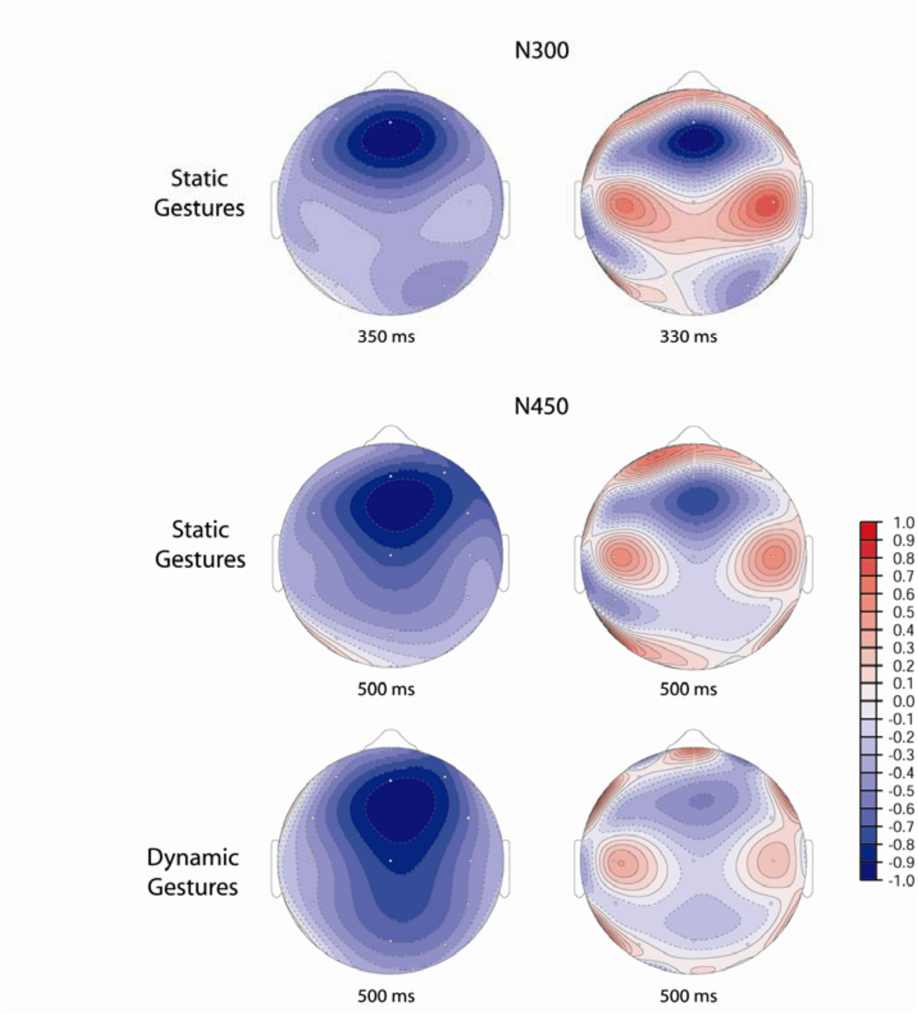


**Figure 5.5.** Normalized isovoltage maps (left) and current source density maps (right) computed from congruency effects elicited by dynamic and static gestures at the peak value of the congruency effect within the N300 (top row) and N450 time windows (bottom 2 rows) post-stimulus onset. (Activity elicited by dynamic gestures is not shown for the N300 time window because congruency effects do not start until 400 ms post-stimulus.)

As shown in the isovoltage maps, both the N300 and N450 exhibit similar scalp topographies, though the N450 is more broadly distributed. Further, the current source density maps suggest that active dipoles within the N300 time window for static gestures are similar to those active during the N450 time window for both gesture types. These views of the data suggest that similar underlying neural generators are engaged both by static and dynamic gestures. Moreover, the N300 effect elicited by static gestures appears to engage similar generators to those mediating the N450 effect. Given these commonalities, it is possible that the N300 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. Nevertheless, similarities in morphology (see Figure 5.2), distribution, and current source density between dynamic and static N450 effects suggest that both categories of stimuli likely engaged a highly similar set of neural generators. Because static gesture stills were specifically extracted at points of maximal meaningfulness and discernibility, it is possible that early congruency effects were evident with these trials, whereas variation in the uniqueness points of dynamic gestures may have prevented their detection within the N300 time window. Further, understanding static gestures may have required participants to draw more extensive inferences about their meaning, resulting in the more broadly distributed late congruency effect observed in response to static items but not in their dynamic counterparts.

### *Picture ERP effects*

The anterior N300 and N400 effects elicited by photographs of common and unidentifiable objects are in keeping with the original findings of McPherson and Holcomb (1999), as well as an earlier experiment involving line drawings (P. J. Holcomb & W. B. McPherson, 1994). Further, the onset of ERP effects obtained in this experiment was remarkably consistent with the data reported in McPherson and Holcomb (1999). It should be noted, however, that different measurement intervals were used to evaluate the N300. While McPherson and Holcomb applied a time window spanning 225 to 325 ms after stimulus onset, the present study assessed measurements obtained between 300 and 400 ms post-stimulus onset in order to maintain consistency with the analysis of gesture trials. Because averaged ERP waveforms elicited during our own presentation of three target types begin to diverge from one another between 100 and 200 ms after stimulus presentation, it is likely that neural generators contributing to the N300 were active well before 300 ms in the present experiment as well, manifesting a time course similar to that reported by McPherson and Holcomb.

However, there are also obvious differences between the outcomes of the current and previous experiments. Notably, brain response to unrelated targets remained consistently more negative than related ones through the end of the epoch, whereas the unrelated and unidentifiable targets did not reliably differ. By contrast, McPherson and Holcomb (1999) report that after 600 ms post-stimulus, reliable differences were detectable in the case of unidentifiable and unrelated images, but not related and unrelated ones.

One likely factor contributing to these discrepancies is the considerably longer duration of time for which targets remained on the computer monitor in the present experiment (2.3 seconds) relative to 400 ms in McPherson and Holcomb's study. Presumably, this longer duration engendered more extensive semantic processing of unrelated items, thereby increasing differences between unrelated and related trials, and concomitantly reducing differences between unrelated and unidentifiable ones. Disparities between outcomes observed here and in previous studies may also be task-driven. Both McPherson and Holcomb (1999) and Holcomb and McPherson (1994), for example, required participants to overtly classify targets according to binary dimension of either relatedness or recognizability, whereas participants in our study were instructed simply to attend to all picture stimuli.

### *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 congruency effects measured in the same time windows. 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 N300 and N400 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 5.6).

Because current source density maps are more sensitive the configuration of current flow, CSD maps (Figure 5.7) were computed from the isovoltage maps depicted in Figure 5.6. Comparing within stimulus types, Figure 5.7 indicates a fairly stable spatial pattern of sources and sinks at the peak of N300 and N400 effects. However, comparing between stimulus types reveals largely heterogeneous distributions of current flow. This outcome suggests that while static gestures elicit N300 and N400 responses that are similar in distribution and functional characterization to those elicited by photographs of common objects, these responses are likely mediated by possibly overlapping, but non-identical neural generators.

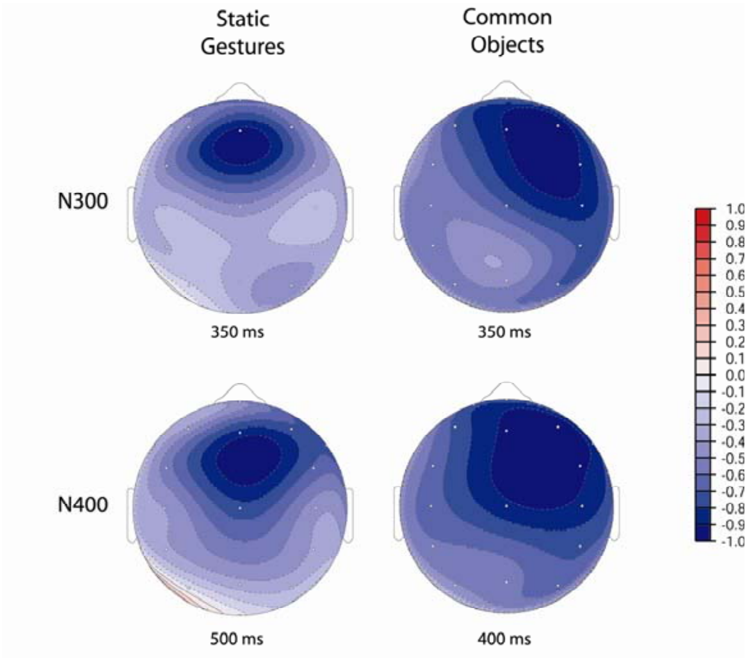


**Figure 5.6.** Normalized isovoltage maps of N300 and N400 effects elicited by static gestures and object photos. Because the object photo N400 peaks earlier than the static gesture N450, these two effects are plotted at different time points.

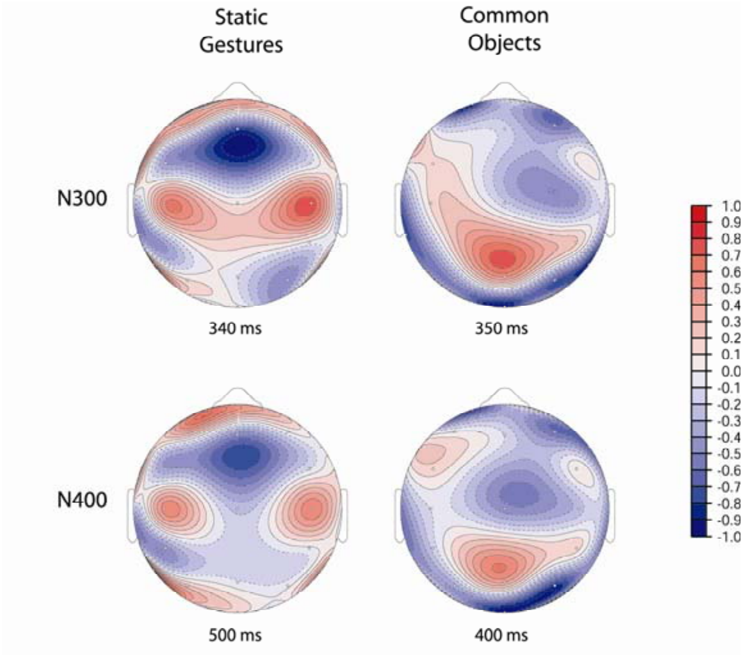


**Figure 5.7.** Current source density constructed from isovoltage maps of the N300 and N400 congruency effects elicited by static gestures (left) and the relatedness effects elicited by photographs of common objects.

This proposal is hardly surprising, given that neuroimaging research has implicated distinct brain systems in the visual processing of objects and human bodies. Viewing images of the face and body, but not inanimate objects, has been shown to engage cortical regions including the extrastriate body area (EBA) and the fusiform face area (FFA) (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006) , as well as the superior temporal sulcus (STS) (Allison, Puce, & McCarthy, 2000). On the other hand, category specific activations have been reported in ventral and lateral portions of the posterior temporal lobe in response to distinct classes of objects, such as animals and tools (Chao, Haxby, & Martin, 1999). Further, 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 fundamentally differ from photographic representations not only in their physical instantiation through the human body, but also in the kinds of information that they convey. On the one hand, 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. On the other, 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 iconic 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, iconic 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 resemble a wick in any way, it can nevertheless be construed as part of the falling candle.

A second example of analogy-based representation in iconic gestures is illustrated on a different occasion, when the same speaker describes placing two slices of bread side by side in order to make a peanut butter sandwich, and simultaneously holds out his hands in front of him, also side by side, with palms facing upward. Significantly, turning his extended palms upwards involves more rotation of the forearm than holding them in a downward facing position. However, because the palm is the inner side of the hand, it can be used analogically to represent the inner side of the sandwich bread (that is, the side on which peanut butter will be spread). By contrast, a simple drawing or photograph of a slice of bread does not allow for the representation of the inner side of the sandwich in progress.

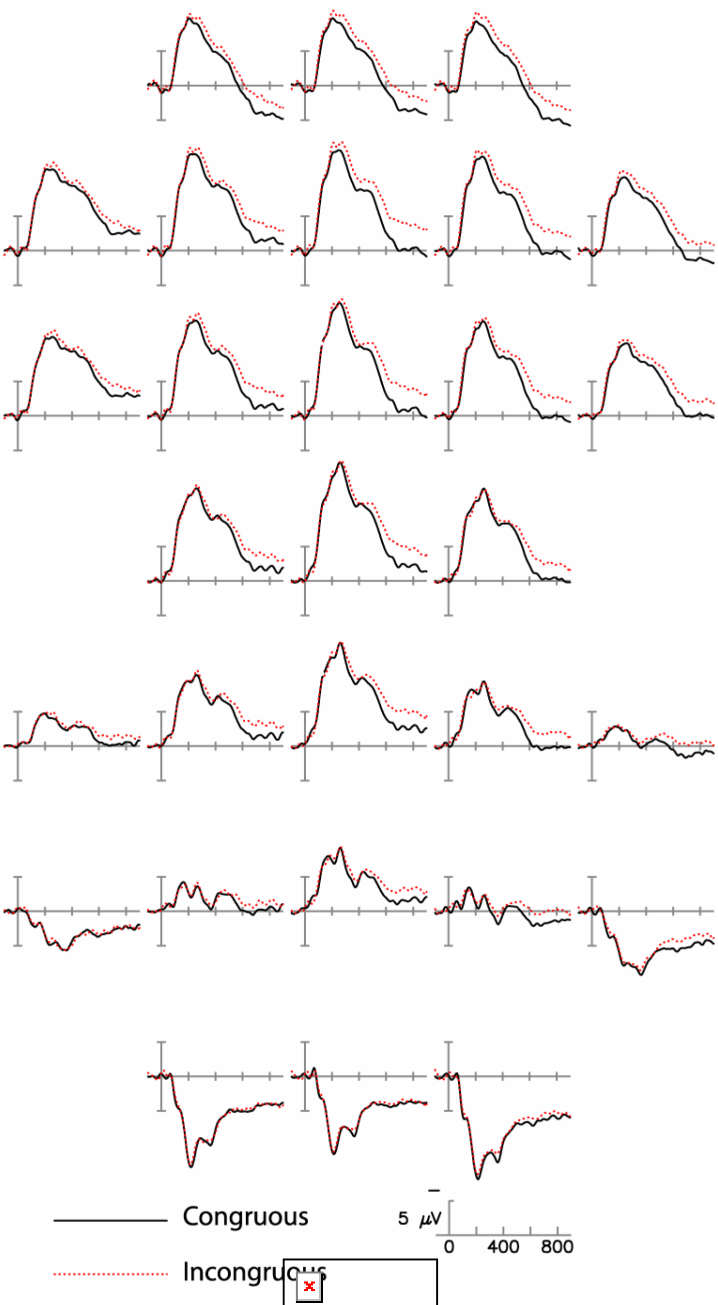
Given these differences in the representational capacities of gestures and pictures, it is not surprising that they elicit somewhat different patterns of neural activation. However, the functional and topographical similarities between N300 and N400 effects elicited by these stimulus types suggest possible commonalities between the systems mediating gesture and picture comprehension. Further analysis of these data with non-linear source separation techniques, such as independent components analysis, may help to resolve this question.

## Conclusion

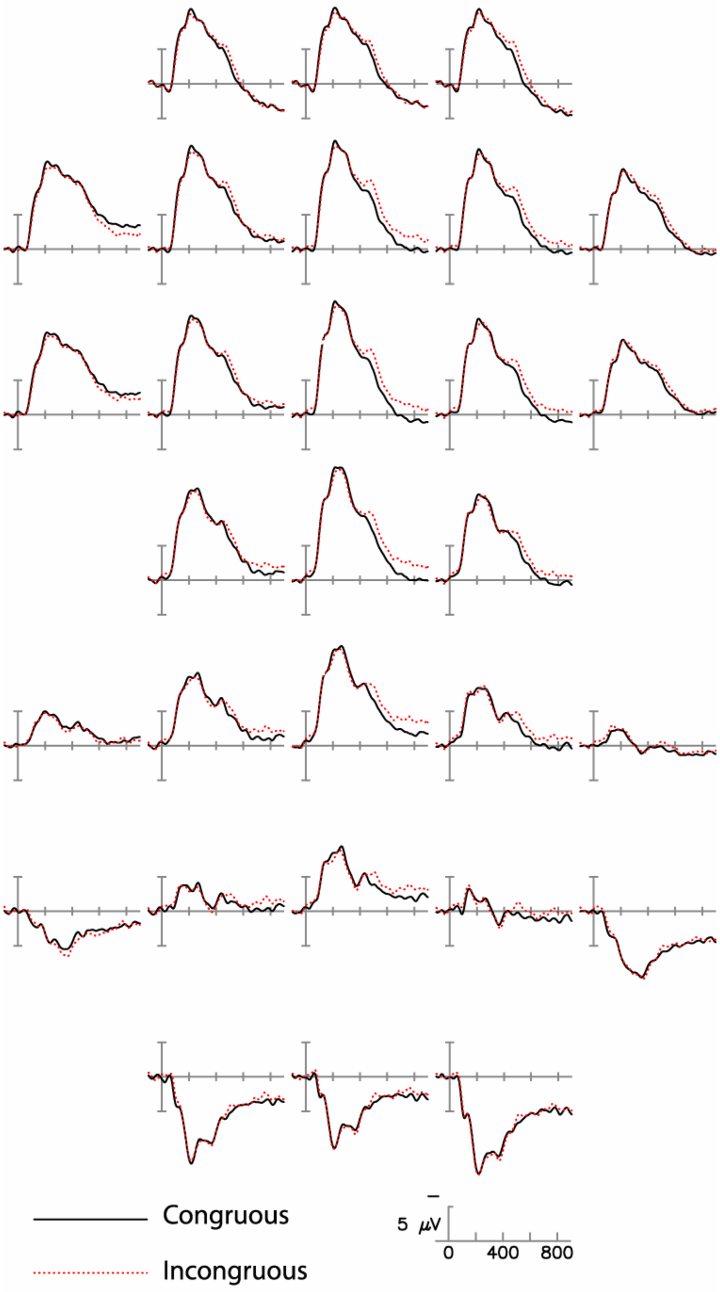
The present study has confirmed and advanced existing research on the on-line comprehension of gestures. In keeping with previous findings (Wu & Coulson, 2004), ERPs time locked to the onset of contextually congruent and incongruent iconic co-speech gestures elicited a broadly distributed negative component – the N450 – with enhanced amplitude for incongruous as compared to congruous 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 percept activated by static gesture stills is affected by the representational properties of the gestures, such as hand configurations and body posture. In this way, iconic hand configurations are similar to contours and shapes that allow pictures of objects to be successfully recognized.

On the other hand, however, N300 and N450 effects elicited by static gestures engaged a different underlying set of neural generators from those responsive to photographs of common objects, as illustrated by CSD maps of congruency effects. This finding is consistent with the different representational properties of pictures and iconic gestures.

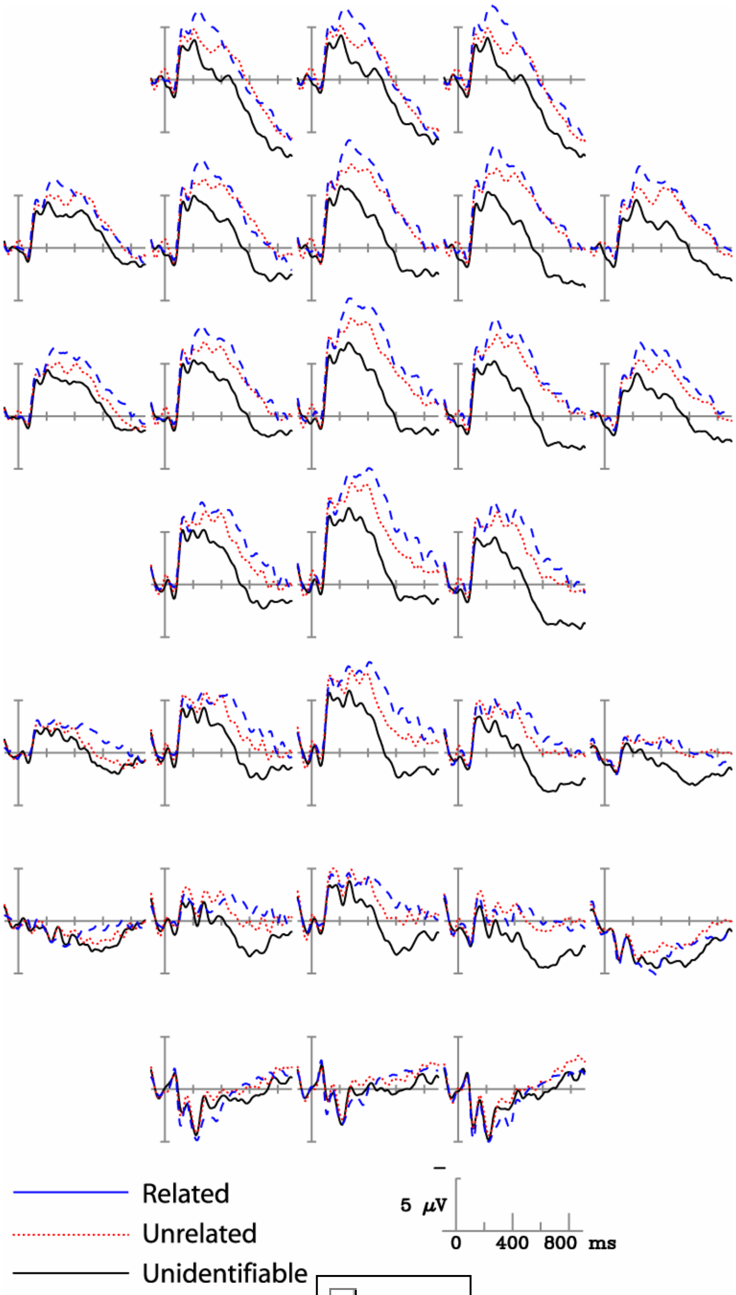
## Appendix



ERPs time-locked to the onset of congruous and incongruous static gestures



ERPs time-locked to the onset of congruous and incongruous dynamic gestures.



ERPs elicited by related, unrelated, and unidentifiable picture probes

## Works Cited

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.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS. *Trends in Cognitive Sciences, 4*(7), 267-278.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*(4), 600-609.

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

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *developmental Psychobiology, 40*(3), 293-310.

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., et al. (1998). Hemispheric specialization for English and ASL: left invariance-right variability. *Neuroreport, 9*(7), 1537-1542.

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*(2), 179-192.

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language, 31*(2), 308-327.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas ina somatopic manner: an fMRI study. *European Journal of Neuroscience, 13*, 400-404.

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*(6), 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), 1-33.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience, 2*(10), 913-919.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism, 7*(4), 303-319.

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), 1-18.

Corina, D. P., Poizner, H., Bellugi, U., Feinberg, T., Dowd, D., & O'Grady-Batch, L. (1992). Dissociation between linguistic and nonlinguistic gestural systems: A case for compositionality. *Brain and Language, 43*(3), 414-447.

Coulson, S., & Federmeier, K. D. (2003). Words in context: ERPs and the lexical/postlexical distinction. *Journal of Psycholinguistic Research.*

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

Deacon, D., Hewitt, S., Yang, C. M., & Nagata, M. (2000). Event-related portential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research, 9*, 137-146.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain, 120*, 1763-1777.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilitstic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience, 8*(8), 117-1121.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16*(10), 1453-1461.

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

Emmorey, K. (1999). Do signers gesture? In R. Campbell (Ed.), *Gesture, speech, and sign*. Oxford: Oxford University Press.

Emmorey, K., & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture, 1*(1), 35-50.

Fadiga, L., Fogassi, G., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology, 73*(2608-2611).

Fauconnier, G. (1985). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge: 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. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(202-224).

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

Feyereisen, P., & de Lannoy, J. D. (1991). *Gestures and speech: Psychological investigations*. New York, NY: Cambridge University Press.

Feyereisen, P., van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech. *European Bulletin of Cognitive Psychology, 8*(3-25).

Gallagher, H., & Frith, C. D. (2004). Dissociable neural pathways for the perception and recogniton of expressive and instrumental gesture. *Neuropsychologia, 42*(13), 1725-1736.

Gallese, V., Fadiga, L., Fogassi, G., & Rizzolatti, G. (1996). Action recogntion in the premotor cortex. *Brain, 119*, 593-609.

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*(2), 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), 1-26.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, Massachusetts: Belknap Press.

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

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge thorugh gestures: Using children's hands to read their minds. *Cognition and Instruction, 9*(3), 201-219.

Goodwin, C. (2000). Action and Embodiment within Situated Human Interaction. *Journal of Pragmatics, 32*(10), 1489-1522.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research, 112*, 103-111.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology, 10*, 57-67.

Grezes, J., Costes, N., & Decety, J. (1998). Top-down effect of strategy on the perception of human biological motion: a PET investigation. *Cognitve Neuropsycholgoy, 15*(6/7/8), 553-582.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping, 12*, 1-19.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*, 1409-1422.

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

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture: Implications for gesture classification and function. *Journal of Language and Social Psychology, 23*(2), 204-214.

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*(2), 149-169.

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

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 fo rthe 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.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281-312.

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: AI Press Inc.

Johnson-Frey, S. H., Maloof, M. R., Newman-Norland, R., Farrer, C., Inati, S., & Grafton, S. T. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neruon, 39*, 1053-1058.

Kellenbach, M. L., Wijers, A. A., & Mulder, G. (2000). Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Cognitive Brain Research, 10*, 67-75.

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language, 28*, 325-349.

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., & Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*(1), 85-93.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2001). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 243-260.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 243-260.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177-210). New York: Pergamon Press.

Kendon, A. (2000). Language and gesture: unity or duality? In D. McNeill (Ed.), *Language and Gesture* (pp. 47-62). UK: Cambridge U Press.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: 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*(5), 743-754.

Kutas, M., Federmeier, K., Coulson, S., King, J. W., & Muente, T. F. (2000). Language. In G. G. Berntson (Ed.), *Handbook of Psychophysiology* (2nd ed., pp. 576-601). Cambridge, UK: Cambridge University Press.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12).

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

Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203-305.

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.

Kutas, M., Van Petten, C. K., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*(3), 218-233.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

LeBaron, C., & Streeck, J. (2000). Gestures, Knowledge, and the World. In D. McNeill (Ed.), *Language and Gesture* (pp. 118-138). UK: Cambridge University Press.

Lotze, M., Heymans, U., Birbaumer, N., Veit, R., Erb, M., Flor, H., et al. (2006). Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia, 44*, 1787-1795.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616-618.

MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P. K., et al. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *NeuroImage, 22*, 1605-1618.

Marshall, J., Atkinson, J., Smulovitch, E., Thacker, A., & Woll, B. (2004). Aphasia in a user of British Sign Language: Dissociations between sign and gesture. *Cognitive Neuropsychology, 21*(5), 537-554.

Mayberry, R. I., & Jacques, J. (2000). Gesture production during stuttered speech. In D. McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge University Press.

McCarthy, G., & Wood, C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62*, 203-308.

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

McNeill, D. (1998). Speech and gesture integration. In J. Iverson & S. Goldin-Meadow (Eds.), *The nature and functions of gesture in children's communication. New directions for child development.* (Vol. 79, pp. 11-27). San Francisco: Jossey-Bass.

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

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

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

Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language, 19*(3), 559-580.

Murphy, K. M. (2004). Imagination as joint activity: The case of architectural interaction. *Mind, Culture, and Activity, 11*(4), 267-278.

Muthukumaraswamy, S., & Johnson, B. W. (2004). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology, 41*, 152-156.

Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. (2004). Mu rhythm modulation during observationof an object-directed grasp. *Cognitive Brain Research, 19*, 195-201.

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating american sign language: Effects of sensory experience and age of acquisition. *Brain and Language, 57*(3), 285-308.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15-22.

Nunez, R. E., & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in crosslinguistic comparison of spatial construals of time. *Cognitive Science, 30*(3), 401-450.

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.

Pecher, D., Zeelenberg, R., & Raajimakers, J., G. (1998). Does pizza prime coin? Perceptual priming in lexical decision and pronunciation. *Journal of Memory & Language, 38*, 401-418.

Peigneux, P., Van der Linden, M., Garraux, G., Laureys, S., Degueldre, C., Aerts, J., et al. (2004). Imaging a cognitive model of apraxia: The neural substrate of gesture-specific cognitive processes. *Human Brain Mapping, 21*, 119-142.

Petrides, M., & Pandya, D. N. (1997). Comparative architectonic analysis of the human and the macaque frontal cortex. In J. Grafman (Ed.), *Handbook of Neuropsychology* (Vol. 9, pp. 17-58). New York: Elsevier.

Pineda, J. A. (2005). The functional significance of mu rhythms: Translating "seeing" and "hearing" into "doing". *Brain Research Review, 50*, 57-68.

Plante, E., Van Petten, C. K., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia, 38*, 1669-1684.

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

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

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychological Bulletin, 89*(3), 506-540.

Rime, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology, 12*(2), 113-129.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*, 188-194.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996). Localization of grasp representations in humans by PET. *Experimental Brain Research, 111*, 246-252.

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

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience, 24*(27), 6181-6188.

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

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience, 15*(1), 111-135.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85-89.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160-164.

Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Review of Neurobiology, 14*(3-4), 199-224.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153-156.

Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology, 42*(1), 144-168.

Thompson, L. A., & Massaro, D. W. (1994). Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology, 57*, 327-354.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology, 39*, 791-808.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teacher's gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology 29*(187-204).

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*(6), 657-671.

Van Petten, C. K., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394-417.

Van Petten, C. K., & Kutas, M. (1990). Interactions between sentence context and word frequency in even-related brain potentials. *Memory & Cognition, 18*(4), 380-393.

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*(3), 363-375.

Wicha, N. Y. Y., Bates, E., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165-168.

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

Wu, Y. C. (submitted). Meaning in gestures: What event-related potentials reveal about processes underlying the comprehension of iconic gestures.

Wu, Y. C., & Coulson, S. (2004). Is that a meaningful gesture: Electrophysiological indices of gesture comprehension. *A supplement to the Journal of Cognitive Neuroscience*, 31.

Wu, Y. C., & Coulson, S. (2005). Is that a meaningful gesture: Electrophysiological indices of gesture comprehension. *Psychophysiology, 42*(6), 654-667.

Wu, Y. C., & Coulson, S. (in press). Is that a meaningful gesture: Electrophysiological indices of gesture comprehension. *Psychophysiology*.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168-171.

# CHAPTER 6: GENERAL DISCUSSION

This research endeavor began with an example of an individual using both language and gesture to describe narrowly avoiding a rattlesnake beneath his foot. Over the course of the subsequent chapters, we explored several questions designed to assess the impact of gestures such as these on listener comprehension. Through lexical and pictorial priming paradigms, we sought to evaluate semantic activations prompted both by gestures alone, and by discourse segments involving gesture and speech together. Additionally, we sought to characterize semantic processes mediating gesture comprehension. How are iconic gestures integrated with other contextually activated information? Does understanding the representational content of iconic gestures involve object recognition processes similar to those engaged by actual objects?

This compilation of studies is motivated by the overall goal of operationalizing theoretical claims regarding co-speech gesture comprehension in terms of basic cognitive and neural function. One influential idea that has inspired a number of behavioral experiments is the proposal that information from speech and gesture is synthesized in a unified underlying representation by listeners. The present work delves into cognitive processes that could support such integrated representations. In particular, research described here suggests that listeners understand discourse describing material objects and events by means of image-based semantic activations that are driven both by the speaker's utterances and his gestures.

A second prominent theme in the field of gesture research is the idea that the information made available through iconic gestures complements, but fundamentally differs from that expressed through speech. One salient disparity between these two modes of representation concerns the semiotic properties that render them interpretable. Iconic gestures, as their name suggests, have been characterized as depictive and mimetic (Goldin-Meadow, 2003; McNeill, 1992, 2005), whereas the language used in speech derives its capacity for signification from largely arbitrary associations between sound and meaning.

This semiotically based distinction between language and gesture has also led researchers to remark on structural differences as well. McNeill (1992; 2005) points out that language encodes semantic content through segmentable, hierarchically organized units, whereas gestures do so globally – that is, the meaning of the gesture as a holistic unit determines the meaning of its individually discernible parts. Further, in the case of language, individual units usually correspond to distinct elements of meaning, whereas a number of semantic relations can be simultaneously expressed in a single gesture.

Although the present set of studies does not directly compare speech and gesture comprehension, it brings to light both commonalities and differences in the cognitive systems that they engage. As shown in Chapter 2 and 3, gestures induce semantic integration and priming processes analogous to those triggered by contentful language. However, as shown in Chapter 5, information in iconic gestures also appears to affect processes specific to the comprehension of visual representations as well. These findings are consistent with the view that unlike words, the communicative function of iconic gestures derives crucially from their depictive properties. However, they also suggest that in spite of the structural and semiotic differences noted above, language and gesture may not involve radically different comprehension systems.

### ***Speech-gesture integration***

The idea that speech and gesture are integrated during comprehension has found support in a number of behavioral studies. A particularly important finding with respect to this issue was obtained by presenting viewers with videos of an actor, who, in the course of describing a cartoon, sometimes produced gestures that deliberately conflicted with information in his speech (Cassell, McNeill, & McCullough, 1999; McNeill, 1992). When viewers were subsequently prompted to retell what they had understood from the video, it was discovered that they expressed information that had been unique to the actor's gestures through speech more often than through their own gestures.

This outcome was hailed as important evidence in favor of the speech-gesture integration hypothesis. It demonstrated that information in gesture is available for subsequent recoding in linguistic form, as one would expect if the semantic content of speech and gesture is integrated in a uniform underlying representation. However, this finding does not preclude alternative interpretations. It is possible that information from the actor's speech and gestures was not actually *integrated* in the sense that listeners formed coherent mental models on the basis of input from each modality. Given the fact that many of the stimulus gestures directly conflicted with their accompanying speech, it may be the case that meanings activated by each modality tended simply to be encoded, but not reconciled. The dependent measures utilized by the authors – namely, the quantity and modality (verbal or gestural) of discrepancies produced by experimental participants while retelling the actor's original narration – cannot resolve this question.

Thus, the study conducted by Cassell et al (1999) provides evidence that listeners are sensitive to the semantic properties of a speaker's gestures. Further, it shows that listeners are able to communicate information gleaned from gestures by means of structurally and semiotically distinct systems of representation, such as language. However, this study does not provide insight into the conjoint effects of gesture and speech on the underlying conceptual processes engaged during comprehension.

The present work addresses this shortcoming. We begin from the view that speech-gesture integration involves a number of interrelated subprocesses that mediate the decoding of the incoming visual and auditory signal, the activation and selection of meaning, and contextual integration. By measuring brain response to picture probes, the study described in Chapter 4 uncovered two important clues about meaning selection when speakers describe the physical world through concurrent utterances and gestures processes. First, this type of discourse prompted listeners to formulate image-based representations of the item or event being described. Secondly, these representations incorporated distinct, but complementary meanings expressed propositionally through language and analogically through gesture. These findings demonstrate that listeners not only encode information from iconic gestures, but also use this information to structure their conceptual models of the speaker's message.

A number of intriguing questions have also emerged from this study. Further research is needed, for example, to assess the impact of different kinds of iconic gesture meanings on comprehension. In the discourse primes used in our study, gestures usually added information about some combination of visuo-spatial properties, including shape (e.g. a round rather than rectangular rug), orientation (e.g. the angle at which a cell phone is slanted), relative location (e.g. two throw pillows arranged on opposite ends of a couch), and size (a small hourglass timer as opposed to a large one). If listeners benefit to differing degrees from gestures expressing distinct types of visuo-spatial cues, this differential sensitivity should be detectable in brain response to picture probes.

Another question that could be addressed through follow-up research is the degree of independence between gesture- and speech-driven semantic activations. One possibility is that each modality of input accesses distinct, but partially overlapping fields of knowledge in long-term memory. In this view, speech is expected to activate category-specific information about the object or event being described, while a gesture is more likely to weakly activate many different categories of items whose visuo-spatial properties are congruent with the structure of the gesture. Those items that receive overlapping activation from both channels would ultimately be selected for assimilation into the listener's ongoing discourse model. An alternative possibility, however, is that information activated by speech and gestures is largely complementary. In this case, gestures would serve to selectively amplify certain elements within the broader set of category-specific activations driven by speech.

It would be possible to adjudicate between these two juxtaposed views by adding to the picture priming paradigm used in Chapter 4 an additional condition involving picture probes that are related to information expressed exclusively in the speaker's gestures, and not his speech (Gesture-only Related). If gesture-only probes elicit less N300 or N400 when preceded by related discourse primes relative to unrelated ones, we could infer that gesture-based semantic activations are not contingent upon those driven by speech. On the other hand, the finding that ERPs elicited by related gesture-only probes do not reliably differ from those elicited by unrelated items would suggest that speech context constrains the range of semantic activations triggered by gestures (although null results should always be interpreted with caution). Outcomes of this proposed experiment would elucidate whether the comprehension of gestures and speech should be viewed as separate, but interrelated processes, or whether gesture comprehension should be treated essentially as a sub-process that is regulated by the comprehension of ongoing speech.

### *Language and gesture*

The present body of research has revealed both commonalities and differences in the cognitive systems mediating language and gesture comprehension. One important similarity described in Chapter 2 regards the contextual integration of information activated by words and gestures. By analogy to words, which elicit an N400 component modulated by the degree of fit between the stimulus and preceding context, we found that contextually incongruous gestures elicit more negative N450 than congruous ones. That incongruent words and gestures elicit functionally similar brain responses suggests that both modes of representation cause stored knowledge to become active during comprehension, and that comprehenders use this information to update their current situation models.

A second parallel between word and gesture comprehension concerns the capacity of both symbol types to activate semantic information in the absence of supporting context. Some researchers have claimed that the meaning of iconic gestures is driven largely by the speech with which they are co-ordinated. However, in Chapter 3, we explored the hypothesis that even when presented without accompanying speech, gestures activate information related to referents whose visuo-spatial properties are consistent with the gestures' features. If these activations are sufficient to facilitate word comprehension, we would expect related probe words to be easier to comprehend than unrelated ones. This prediction was confirmed in two experiments, wherein probe words elicited less N400 in cases when they were related to a preceding gesture than in cases when they were not.

On the other hand, the present research also allows differences to be posited between word and gesture comprehension. Most notably, iconic gestures engage systems responsive to image-based rather than lexical representations. In Chapter 5, we compared ERPs elicited by photographs of common objects and "static gestures," which were created by extracting from each dynamic gesture movie clip a single freeze frame that made visible critical information about the speaker's intended meaning. Both static gestures and object photos yielded similar N300 congruency effects, suggesting that at least the static semiotic elements of gestures, such as hand shape, location, and orientation, engage object recognition processes in a manner analogous to the basic structural features of a picture or line drawing. However, differences in the distribution of ERP effects, and local sources and sinks of radial current elicited by gestures and pictures suggested non-identical underlying neural generators. These findings suggest that "seeing" objects represented in gesture does not involve the same mechanisms recruited during the identification of actual objects.

A second difference regards the kinds of information activated by language and gesture. In Chapter 4, it was found that picture probes that were cross-modally related to information in both speech and gestures elicited less N300 than probes related to speech alone. This result was interpreted as evidence that the gestures occurring in discourse primes activated image-specific information enabling participants to recognize cross-modally related probes more easily than speech-only ones.

Given our findings, it is tempting to postulate further that gestures trigger image-based semantic activations that are not induced by language. However, there is also evidence that concrete words may affect comprehension in a manner parallel to image-based representations. In particular, it has been shown that by comparison with abstract words, concrete ones elicit more negative ERPs over anterior electrode sites within the time window of the N400 (Holcomb, Kounios, Andersen, & West, 1999; Kounios & Holcomb, 1994). This effect is thought to reflect the access and integration of imagistic information associated with concrete, but not abstract words. Because the anterior distribution of this effect is consistent with anterior N300 and N400 effects elicited by photographs, line drawings, and in the present research, iconic gestures, it is possible that concrete words, gestures, and pictures engage overlapping neural resource, and may therefore impact conceptual activity in similar ways.

To summarize, the present body of research suggests commonalities in the systems devoted to language and gesture comprehension at the level of conceptualization. Chapter 2 provides support for the idea that comprehenders integrate information from iconic gestures with higher-order discourse level representations in a manner analogous to the contextual integration of language-based input. Further, the finding in Chapter 3 that iconic gestures facilitate the comprehension of related words is consistent with the view that both of these modes of representation engage a common underlying conceptual substrate. On the other hand, however, the meanings of words and gestures appear to be accessed by modality–specific mechanisms, as indicated in Chapter 5 by the finding that iconic gestures engage higher-order visual processes implicated in object recognition.

### *Future Directions*

How do individuals decode information represented in iconic gestures? As explored in Chapter 5, one bottom-up approach to this problem builds from the idea that gestures are processed in a manner similar to more conventionally depictive representations, such as photographs or line drawings. This view is intuitively appealing, given the fact that speakers often use iconic gestures to indicate edges, surfaces, and global shapes of objects. It is also important, however, to explore the possibility that different types of iconic gestures may engage distinct recognition systems, or that multiple systems may mediate the decoding process.

One potentially exciting area of research in line with these ideas involves comparing the underlying substrates mediating the comprehension of action and gesture. This approach is promising, given the accumulation of evidence over the past decade that understanding and executing actions involves overlapping neural systems. A cornerstone of the motor theory of action comprehension is the discovery of mirror neurons, which are cells found in the rostral part of ventral premotor cortex (F5) in the macaque monkey. They have been shown to discharge both when a macaque grasps and manipulates objects and when it observes another macaque or human experimenter perform a similar action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Subsequent research has uncovered cells that exhibit a similar response profile in area PF (7b), which comprises the rostral portion of the inferior parietal lobule, and which projects to portions of ventral premotor cortex, including area F5. These two regions are proposed to form a mirror-neuron circuit that presumably constitutes an important part of cortical networks that underlie action comprehension in animals.

Converging lines of evidence suggest that a mirror system also exists in humans. For example, motor-evoked potentials (MEPs) recorded from hand muscles were measured as participants viewed an experimenter grasping objects, or the same objects at rest, while receiving pulses of transcranial magnetic stimulation (TMS) over motor cortex. It was found that during the observation of grasping, but not object viewing, the amplitude of MEPs increased over hand muscles normally recruited in the genuine act grasping, such as the opponens pollicis (thumb) (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995).

A second line of research suggestive of a human mirror system is the electroencephalographic mu rhythm (~8-13 Hz), which arises from generators in sensorimotor cortex and exhibits functional properties similar to those characterizing mirror neurons. Both executing actions and observing actions executed by others results in the attenuation of mu rhythm (for review, see Pineda (2005)). Further, both mirror neurons and mu rhythm exhibit enhanced sensitivity to goal-directed hand actions as compared with hand movements which do not effect immediately discernible outcomes (Johnson-Frey et al., 2003; S. Muthukumaraswamy & Johnson, 2004; S. D. Muthukumaraswamy, Johnson, & McNair, 2004). Given these similarities in response profile, mu suppression has been proposed to reflect downstream modulation of sensorimotor cortex by the mirror system.

Neuroimaging research seeking to localize the mirror system in humans has revealed a distributed cortical network responsive to the observation of meaningful actions, including regions typically implicated in motor function, such as inferior parietal regions thought to mediate sensorimotor integration (see Grezes and Decety (2001) for review), as well as the inferior frontal gyrus (*pars opercularis* BA 44, and *pars triangularis* BA 45; (Buccino et al., 2001; Decety et al., 1997; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Grezes, Costes, & Decety, 1998; Peigneux et al., 2004; Rizzolatti, Fadiga, Matelli et al., 1996; Saygin, Wilson, Hagler, Bates, & Sereno, 2004). Because the *pars opercularis* is considered the human homologue to area F5 in the macaque (Petrides & Pandya, 1997), it is highly likely that this region comprises part of a mirror-based action recognition system in humans.

By analogy to actions, the representational content of a speaker's gestures may be accessible to the listener because they activate certain types of reciprocal motoric responses within the listener's own action and gesture production system. This idea is consistent with the proposal espoused by some discourse analysts (LeBaron & Streeck, 2000) that gestures depicting visuo-spatial properties of objects may derive their capacity for signification from links to actions typically performed upon those objects rather than from perceptual similarities to them. To apply this notion to the coiled snake gesture discussed in the General Introduction, for example, the circular movement produced by the speaker with his extended arm and index finger might be evocative of the kind of pointing behavior that would be produced if an individual were drawing attention to an actual object on the ground with a small, circular shape. Likewise, in the case of the double door gesture pictured in Chapter 4 (Figure 4.1), the speaker's flat, sidewaysturned palms might evoke the hand configuration used to palpate a door's flat, vertical surface.

The proposal that iconic gestures may be interpretable because they engage mirror-based action processing systems is also consistent with growing evidence that portions of motor cortex, including the inferior frontal gyrus, are active not only during the observation of actions, but also certain types of gestures. Action pantomimes (e.g., opening a bottle) have activated this region relative to static hands (Grezes et al., 1998) and ASL hand signs whose meaning was not known to participants (Decety et al., 1997). Additionally, emblematic gestures signaling requests (e.g., "Come here") relative to hand actions (Lotze et al., 2006) or relative to a static image of an individual at rest resulted in activation of Broca's area (Gallagher & Frith, 2004).

To summarize, further study is needed to determine the extent to which gestures are processed as actions or objects, or a combination of the two. Although some preliminary evidence in favor of the object recognition view was reported in Chapter 5, it is important to note that this project does not form a fully completed research program. We are currently in the process of identifying signal contributions that are common to the semantic processing of both objects and gestures, as well as those which are unique to each stimulus type by means of independent components analysis (ICA). Further, we hope to conduct additional research to evaluate the functional similarity between N300 and N400 components elicited by object photos and static gestures. As discussed in Chapter 5, for example, unidentifiable objects elicit more N300 and N400 than identifiable ones. By analogy, one might expect meaningless movements to elicit larger N300 and N450 than meaningful gestures. Moreover, just as highly related object photos result in reduced N400 relative to moderately related ones, gestures that are readily interpretable should elicit reduced N450 relative to those that are inherently vague and subject to a wide range of possible construals.

Ultimately, it should be kept in view that the capacity to communicate through gesture derives not only from bottom-up decoding systems – whatever their characteristics may be – but also from the nature of stored knowledge in semantic memory. Numerous converging lines of evidence – including linguistic phenomena, such as classifiers; psychological phenomena, such as the primacy of basic-level categories; and neurobiological phenomena, such as feature detectors – indicate that elements of human experience are represented in highly abstract, schematic forms (see Lakoff (1987) for review). Thus, a speaker's success in conveying something about the shape of a snake through a brief circular sweep of his hand and arm is likely to depend not only upon the listener's ability to identify the visuo-spatial features of this gesture, but also upon the availability of comparably schematic shape representations on the part of the listener.

## Works Cited

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas ina somatopic manner: an fMRI study. *European Journal of Neuroscience, 13*, 400-404.

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), 1-33.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain, 120*, 1763-1777.

Fadiga, L., Fogassi, G., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology, 73*(2608-2611).

Gallagher, H., & Frith, C. D. (2004). Dissociable neural pathways for the perception and recogniton of expressive and instrumental gesture. *Neuropsychologia, 42*(13), 1725-1736.

Gallese, V., Fadiga, L., Fogassi, G., & Rizzolatti, G. (1996). Action recogntion in the premotor cortex. *Brain, 119*, 593-609.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, Massachusetts: Belknap Press.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research, 112*, 103-111.

Grezes, J., Costes, N., & Decety, J. (1998). Top-down effect of strategy on the perception of human biological motion: a PET investigation. *Cognitve Neuropsycholgoy, 15*(6/7/8), 553-582.

Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping, 12*, 1-19.

Holcomb, P. J., Kounios, J., Andersen, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721-742.

Johnson-Frey, S. H., Maloof, M. R., Newman-Norland, R., Farrer, C., Inati, S., & Grafton, S. T. (2003). Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neruon, 39*, 1053-1058.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 804-823.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

LeBaron, C., & Streeck, J. (2000). Gestures, knowledge, and the world. In D. McNeill (Ed.), *Language and Gesture* (2 ed., pp. 118-140). Cambridge: Cambridge University Press.

Lotze, M., Heymans, U., Birbaumer, N., Veit, R., Erb, M., Flor, H., et al. (2006). Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia, 44*, 1787-1795.

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

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.

Muthukumaraswamy, S., & Johnson, B. W. (2004). Changes in rolandic mu rhythm during observation of a precision grip. *Psychophysiology, 41*, 152-156.

Muthukumaraswamy, S. D., Johnson, B. W., & McNair, N. (2004). Mu rhythm modulation during observationof an object-directed grasp. *Cognitive Brain Research, 19*, 195-201.

Peigneux, P., Van der Linden, M., Garraux, G., Laureys, S., Degueldre, C., Aerts, J., et al. (2004). Imaging a cognitive model of apraxia: The neural substrate of gesture-specific cognitive processes. *Human Brain Mapping, 21*, 119-142.

Petrides, M., & Pandya, D. N. (1997). Comparative architectonic analysis of the human and the macaque frontal cortex. In J. Grafman (Ed.), *Handbook of Neuropsychology* (Vol. 9, pp. 17-58). New York: Elsevier.

Pineda, J. A. (2005). The functional significance of mu rhythms: Translating "seeing" and "hearing" into "doing". *Brain Research Review, 50*, 57-68.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996). Localization of grasp representations in humans by PET. *Experimental Brain Research, 111*, 246-252.

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience, 24*(27), 6181-6188.