Gesture's neural language

Michael Andric[1,2]* **and Steven L. Small**[1,2,3]

[1] Department of Psychology, The University of Chicago, Chicago, IL, USA
[2] Brain Circuits Laboratory, University of California Irvine, Irvine, CA, USA
[3] Department of Neurology, University of California Irvine, Irvine, CA, USA

When people talk to each other, they often make arm and hand movements that accompany what they say. These manual movements, called "co-speech gestures," can convey meaning by way of their interaction with the oral message. Another class of manual gestures, called "emblematic gestures" or "emblems," also conveys meaning, but in contrast to co-speech gestures, they can do so directly and independent of speech. There is currently significant interest in the behavioral and biological relationships between action and language. Since co-speech gestures are actions that rely on spoken language, and emblems convey meaning to the effect that they can sometimes substitute for speech, these actions may be important, and potentially informative, examples of language–motor interactions. Researchers have recently been examining how the brain processes these actions. The current results of this work do not yet give a clear understanding of gesture processing at the neural level. For the most part, however, it seems that two complimentary sets of brain areas respond when people see gestures, reflecting their role in disambiguating meaning. These include areas thought to be important for understanding actions and areas ordinarily related to processing language. The shared and distinct responses across these two sets of areas during communication are just beginning to emerge. In this review, we talk about the ways that the brain responds when people see gestures, how these responses relate to brain activity when people process language, and how these might relate in normal, everyday communication.

Keywords: gesture, language, brain, meaning, action understanding, fMRI

INTRODUCTION

People use a variety of movements to communicate. Perhaps most familiar are the movements of the lips, mouth, tongue, and other speech articulators. However, people also perform co-speech gestures. These are arm and hand movements used to express information that accompanies and extends what is said. Behavioral research shows that co-speech gestures contribute meaning to a spoken message (Kendon, 1994; McNeill, 2005; Feyereisen, 2006; Goldin-Meadow, 2006; Hostetter, 2011). Observers integrate these gestures with ongoing speech, possibly in an automatic way (Kelly et al., 2004; Wu and Coulson, 2005). In contrast, people also use what are called emblematic gestures, or *emblems*. These are hand movements that can convey meaning directly, independent of speech (Ekman and Friesen, 1969; Goldin-Meadow, 2003). A familiar example is when someone gives a "thumbs-up" to indicate agreement or a job well done. Emblems characteristically present a conventional visual form that conveys a specific symbolic meaning, similar in effect to saying a short phrase like "Good job!" Still, both co-speech gestures and emblems are fundamentally hand actions. This is important because people encounter many types of hand actions that serve other goals and do not convey any symbolic meaning, e.g., grasping a cup. Thus, in perceiving hand actions, people routinely discern the actions' function and purpose. As this applies to understanding co-speech gestures and emblems, people must register both their manual action information and

their symbolic content. It is not yet clear how the brain reconciles these manual and symbolic features. Recent research on the neurobiology involved in gesture processing implicates a variety of responses. Among these, there are responses that differentially index action and symbolic information processing. However, a characteristic response profile has yet to emerge. In what follows, we review this recent research, assess its findings in the context of the neural processing of actions and symbolic meanings, and discuss their interrelationships. We then evaluate the approaches used in prior work that has examined gesture processing. Finally, we conclude by suggesting directions for future study.

Although they are often considered uniformly, manual gestures can be classified in distinct ways. One way is by whether or not a gesture accompanies speech. Another is by the degree to which a gesture contributes meaning in its own right or in conjunction with speech. That is, manual gestures can differ in the nature of the semantic information they convey and the degree to which they rely on spoken language for their meaning. For example, *deictic* gestures provide referential information, such as when a person points to indicate "over there" and specify a location. Another class, called *beat* gestures, provide rhythm or emphasis by matching downward hand strokes with spoken intonations (McNeill, 1992). Neither deictic nor beat gestures supply semantic information in typical adult communication. In contrast, there are *iconic* and *metaphoric* gestures. These provide semantic meaning

that either complements what is said or provides information that does not otherwise come across in the verbal message (McNeill, 2005). Iconic and metaphoric gestures must be understood in the context of speech. For example, when a person moves their hand in a rolling motion, this can depict wheels turning as an iconic gesture in the context of "The wheels are turning." However, in the metaphoric use of "The meeting went on and on," the same movement can represent prolonged continuation. In other words, the speech that accompanies these gestures is key to their representational meaning.

Because gestures vary in the way they provide meaning, the relation between gestures and language is a complex, but interesting, topic. One view is that gestures and spoken language – at both psychological and biological levels of analysis – share the same communication system and are two complementary expressions of the same thought processes (McNeill, 1992). Many findings support this proposal (Cassell et al., 1999; Kelly et al., 1999; Wu and Coulson, 2005; Bernardis and Gentilucci, 2006). For example, Cassell et al. (1999) found that when people retell a narrative that was presented to them using gestures that do not match the spoken content, their retelling takes into account both the spoken and mismatched gesture information. The relation of gesture to speech is strong enough that their retelling may even include new events that resolve the conflicting speech and gestures. Moreover, another study found that when an actor pointed to an open screen door and said "The flies are out" people were much more likely to correctly understand the intended meaning (here, to close the door) when both speech and gesture were present than if only one or the other was given (Kelly et al., 1999). Thus, the way that people interpret a message is constrained when gestures and speech interact.

What are the neurobiological implications of this view that speech and gestures share a common system? There is, in fact, some neural evidence that gestures may evoke responses in brain areas that are also active when people comprehend semantic information in language. Yet, gestures are hand actions. Thus, it is also important to recognize the neural function associated with perceiving hand actions, regardless of these actions' purpose. In other words, there is a need to reconcile the neurobiology of action understanding with the neurobiology of understanding semantic information.

Prior research (described in detail below) suggests that the neural circuits involved in action understanding primarily include parts of the inferior parietal, premotor, posterior lateral temporal, and inferior frontal cortices. Interestingly, some of these brain areas, particularly in the lateral temporal and inferior frontal cortices, also respond to information conveyed in language. However, it is not known if these responses depend on the modality (e.g., language) by which this information is conveyed. Thus, this prior work leaves a number of open questions. For example, it remains unclear whether brain responses to gestures are primarily driven by the gestures' recognition as hand actions. In other words, it is uncertain whether some brain responses simply reflect sensitivity to perceiving hand actions in general, or if such responses are more tuned to the communicative information that some gestures convey. This would contrast with responses to hand actions that do not directly communicate meaning, such as grasping an object.

Also, as gestures can communicate meaning, it remains to be determined if the meaning they convey is processed in a similar way as when meaning is presented in other forms, such as language. An even more basic issue is that it remains unclear whether there is a typical response profile for gestures, in general.

In the following sections, we first survey the prior research on how the brain processes manual actions, in general (*Relevant brain responses in processing gestures*). In two parts, we next review work on brain responses to gestures that communicate meaning, including emblems and co-speech gestures. We highlight areas that might respond regardless of a hand action's use in communicating meaning (*Perceiving hand actions: Inferior parietal and premotor cortex*). We then focus on brain areas thought to be important for processing meaning in language (*Perceiving meaningful hand actions: Inferior frontal and lateral temporal cortex*).

RELEVANT BRAIN RESPONSES IN PROCESSING GESTURES

People routinely perceive and understand others' hand movements. However, it is not yet clear how the brain processes such information. This is very important for understanding how gestures are recognized, since gestures are fundamentally arm and hand movements. One of the most significant findings to offer insight into a potential neural mechanism of action perception is the discovery of mirror neurons in the macaque brain. These are neurons that characteristically fire both when an animal performs a purposeful action *and* when it sees another do the same or similar act. For example, these neurons fire when the monkey sees an experimenter grasp a piece of food. They stop firing when the food is moved toward the monkey. Then, they fire again when the monkey itself grasps the food. In other words, these neurons fire in response to specific motor acts as each is perceived and performed. Mirror neurons were first found in the macaque premotor area F5 (di Pellegrino et al., 1992) and later in inferior parietal area PF (Fogassi et al., 1998). Given that area F5 receives its main parietal input from anterior PF (Geyer et al., 2000; Schmahmann et al., 2007; Petrides and Pandya, 2009), this circuit is thought to be a "parieto-frontal system that translates sensory information about a particular action into a representation of that act" (Rizzolatti et al., 1996; Fabbri-Destro and Rizzolatti, 2008). This is important because it suggests a possible a neural mechanism that would allow an "immediate, not cognitively mediated, understanding of that motor behavior" (Fabbri-Destro and Rizzolatti, 2008).

The suggestion that a "mirror mechanism" mediates action understanding in monkeys inspired attempts to try to identify a similar mechanism in humans (Rizzolatti et al., 1996, 2002; Rizzolatti and Craighero, 2004; Fabbri-Destro and Rizzolatti, 2008; Rizzolatti and Fabbri-Destro, 2008). This effort began with fMRI studies that examined brain responses when people observed grasping. Results demonstrated significant activity in premotor cortex (Buccino et al., 2001; Grezes et al., 2003; Shmuelof and Zohary, 2005, 2006), as well as parietal areas such as the intraparietal sulcus (IPS; Buccino et al., 2001, 2004; Grezes et al., 2003; Shmuelof and Zohary, 2005, 2006) and inferior parietal lobe. This also includes the supramarginal gyrus (SMG), which is thought to have some homology with monkey area PF (Perani et al., 2001; Buccino et al., 2004; Shmuelof and Zohary, 2005, 2006). For example, Buccino et al. (2001) found that when one person sees another

grasp a cup with their hand, bite an apple with their mouth, and push a pedal with their foot, not only is there parietal and premotor activity, but this activity is somatotopically organized in these areas, similar to the motor cortex homunculus (Buccino et al., 2001). Many studies (Decety et al., 1997; Grezes et al., 2003; Lui et al., 2008; Villarreal et al., 2008) find that these areas also respond when people view pantomimed actions like hammering, cutting, sawing, or using a lighter (Villarreal et al., 2008). This is particularly interesting because, with the object physically absent, it suggests that these areas respond to the action *per se* rather than to the object or to the immediate context. Furthermore, damage to these parietal and premotor areas results in damage to or loss of people's ability to produce and recognize these types of actions (Leiguarda and Marsden, 2000). However, these findings do not clarify whether such responses generalize to hand action observation, or, instead, are specific to observing actions that involve object use. In other words, would these same areas also play a functional role in understanding actions that are used to communicate meaning?
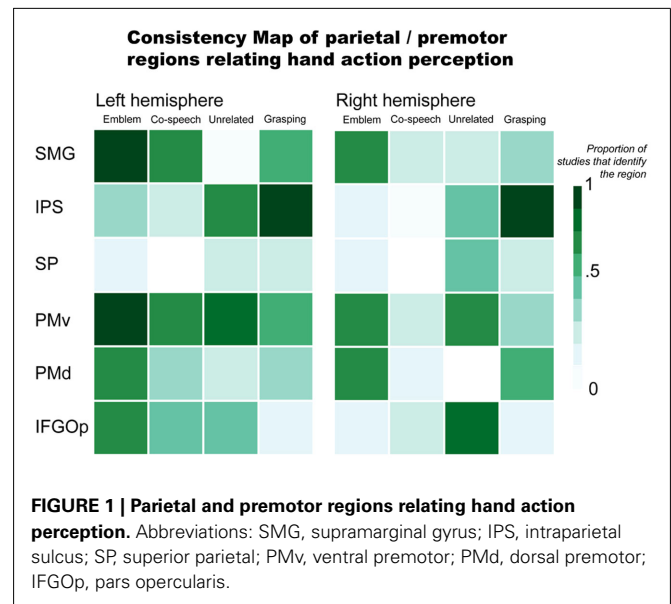
In addressing this question, several authors suggest that such a mirror mechanism might also be the basis for how the brain processes emblems and co-speech gestures (Skipper et al., 2007; Willems et al., 2007; Holle et al., 2008). However, the results needed to support this are not yet established. In particular, it is not clear whether observing a gesture systematically elicits parieto-frontal brain responses. This would be expected if a mirror mechanism based in these areas' function was integral in gesture recognition.

A further unresolved issue concerns whether these meaningful gestures elicit brain responses that are characteristically dissociable from what is found when people see hand actions that are not symbolic, such as grasping an object. These outstanding issues are considered in the following sections.

PERCEIVING HAND ACTIONS: INFERIOR PARIETAL AND PREMOTOR CORTEX

For parietal and premotor regions, their consistent (or inconsistent) reported involvement in gesture processing is illustrated in **Figure 1**. This includes results for processing *emblems*, *co-speech* gestures, hand movements that occur with speech but are *unrelated* to the spoken content, and *grasping* (see Appendix for the list of studies from which data was used to comprise the Figures). These findings most often implicate the ventral premotor cortex (PMv) and SMG as active in perceiving gestures. It is important to remember, however, that gestures can also communicate meaning. To this end, interpreting meaning (most commonly in language) is often linked to brain activity in lateral temporal and inferior frontal regions. Further below, we will address these lateral temporal and inferior frontal regions for their potential roles in gesture processing. Here, we examine parietal and premotor results.

There is evidence that parietal and premotor regions thought to be important in a putative human mirror mechanism respond not just when people view object-directed actions like grasping, but to gestures, as well. Numerous co-speech gesture (Holle et al., 2008, 2010; Dick et al., 2009; Green et al., 2009; Hubbard et al., 2009; Kircher et al., 2009; Skipper et al., 2009) and emblem (Nakamura et al., 2004; Lotze et al., 2006; Montgomery et al., 2007; Villarreal

FIGURE 1 | Parietal and premotor regions relating hand action perception. Abbreviations: SMG, supramarginal gyrus; IPS, intraparietal sulcus; SP, superior parietal; PMv, ventral premotor; PMd, dorsal premotor; IFGOp, pars opercularis.

et al., 2008) studies find inferior parietal lobule activity. More precisely, the SMG and IPS are often implicated. For example, Skipper et al. (2009) found significant SMG responses when people viewed a mix of iconic, deictic, and metaphoric gestures accompanying a spoken story. In this task, the SMG also exhibited strong effective connectivity with premotor cortices (Skipper et al., 2007, 2009). Bilateral SMG activity is found when people view emblems, as well (Nakamura et al., 2004; Montgomery et al., 2007; Villarreal et al., 2008). However, the laterality of SMG effects is not consistent. For example, one study found an effect for the left SMG, but not the right SMG, when people viewed emblems (Lotze et al., 2006). The opposite was found when people saw gestures mismatched with accompanying speech (Green et al., 2009). That is, the effect was identified in the right SMG, not the left.

Both Green et al. (2009) and Willems et al. (2007) suggest that the IPS shows sensitivity when there is incongruence between gestures and speech (e.g., when a person hears "hit" and sees a "writing" gesture). However, these two studies find results in opposite hemispheres: Willems et al. (2007) identify the left IPS and Green et al. (2009) report the right IPS. Right IPS activity is also found to be stronger when people see a person make grooming or scratching movements with the hands ("adaptor movements") than when they see co-speech gestures (Holle et al., 2008). Another study, however, fails to replicate this finding (Dick et al., 2009). The right IPS is also active when people view beat gestures performed without speech (Hubbard et al., 2009). IPS activity is found in emblem studies, as well. But there is again inconsistency across reports. Whereas one study found bilateral IPS activity for processing emblems (Villarreal et al., 2008), others did not report any activity (Lotze et al., 2006; Montgomery et al., 2007). This lack of consistent IPS activity in results for co-speech and emblematic gesture processing is in contrast with results for grasping. That is, results for grasping observation consistently implicate this area. This suggests that the IPS might not play a strong role in interpreting an action's represented meaning *per se*. Rather, IPS responses

may be more tuned in processing a hand action's visuomotor properties. That is, when the focus of the presented information is the hand action, itself, (e.g., in observing grasping or beat gestures without accompanying speech) the IPS responds prominently. This would be the case also when a gesture is incongruent with accompanying speech. In this scenario, as an observer tries to reconcile divergent spoken and manual information, a more detailed processing of the hand action may be required. In contrast, when speech and gestures are congruent, processing the represented meaning, rather than the features of its expression, may be the observer's focus. In such a situation, IPS responses may not be as strong as those of other regions that are more particularly tuned toward interpreting meaning.

Premotor areas are also active when people view gestures. A number of studies report significant bilateral premotor responses to emblems (Nakamura et al., 2004; Montgomery et al., 2007; Villarreal et al., 2008). Some evidence also suggests similar premotor activity for co-speech gesture observation. For example, there is significant bilateral PMv activity when people view metaphoric gestures compared to when they view a fixation cross (Kircher et al., 2009), as well as bilateral PMd activity when they view beat gestures compared to when they watch a still body (Hubbard et al., 2009). These ventral and dorsal distinctions are also found in other studies. Specifically, whereas one study found PMd activity for emblem observation (Villarreal et al., 2008), another found activity localized to PMv, bordering the part of the inferior frontal gyrus (IFG) that also shows sensitivity to emblems (Lotze et al., 2006).

Some research suggests that premotor responses are sensitive to the semantic contribution of gesture. Willems et al. (2007) found left PMv activity to be modulated by the semantic congruency between gestures and speech: left PMv responses were stronger to gestures that were unrelated to what was said compared to when they were congruent. Another study found a similar result with gestures incongruent with a spoken homonym. However, this result implicated *both* left and right PMv cortex (Holle et al., 2008). Finally, Skipper et al. (2009) found that the BOLD signal from bilateral PMv showed a systematic response when people viewed iconic, deictic, and metaphoric gestures during audiovisual story comprehension.

Overall, these findings indicate that parietal and premotor regions are generally active when people view gestures, both as they accompany speech (co-speech gestures) and when they convey meaning without speech (emblems). Yet, parietal and premotor areas do not regularly respond in a way that indicates they are tuned specifically to whether or not the gestures convey meaning. Three primary lines of reasoning support this conclusion: (1) These areas are similarly active when people view non-symbolic actions like grasping as when they view meaningful gestures; (2) Responses in these areas do not appear to systematically distinguish between emblems and co-speech gestures, even though the former directly communicate meaning and the latter rely on speech; and (3) While some findings indicate stronger responses when a gesture does not match accompanying speech than when it does, such findings are not consistent across reports. It seems more likely that these parietal and premotor areas function more generally. That is, their responses may be evident when people view *any*

purposeful hand action, rather than a specific type of hand action. In contrast, areas responsive to the meaning conveyed by these actions are more likely those thought to relate to language understanding, i.e., areas of the inferior frontal and lateral temporal cortices.
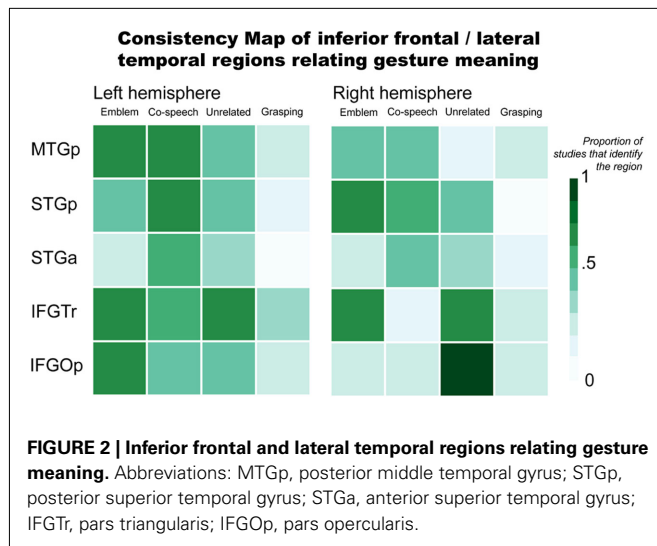
PERCEIVING MEANINGFUL HAND ACTIONS: INFERIOR FRONTAL AND LATERAL TEMPORAL CORTEX

When people see co-speech gestures with associated speech, the gestures contribute to the message's meaning and how it is interpreted (McNeill, 1992, 2005; Kendon, 1994; Goldin-Meadow, 2006). Emblems also communicate meaning. However, in contrast to co-speech gestures, they can do so directly, independent of speech. Emblems can even sometimes be used to substitute for speech (Goldin-Meadow, 1999). In fact, emblems also elicit event-related potentials comparable to those found for words (Gunter and Bach, 2004). Only recently, however, have researchers started studying how the brain processes these gestures. So far, the literature suggests both overlap and inconsistency as to which brain regions are particularly important for their processing. The variation in reported findings may be due to numerous possible sources. For example, one source may be the differing paradigms and analysis methodologies used to derive results. Another may be the extent of results that are given exposition. Also, the way that people interpret these actions may, itself, be highly variable at the neural level. Here, we highlight both the overlapping and varying findings in the gesture literature for regions in the inferior frontal and lateral temporal cortices that may be prominent in processing symbolic meaning.

Brain areas typically associated with language function also respond when people perceive gestures. This is consistent with psychological theories of gesture that propose gesture and language are two ways of expression by a single communication system (McNeill, 1992). In **Figure 2**, we illustrate the consistency of reported inferior frontal and lateral temporal region involvement in gesture processing. This again includes results for *emblems*, *co-speech* gestures, hand movements that occur with speech but are *unrelated* to the spoken content, and, for consistency with the previous figure, *grasping* (see Appendix for the studies that were used to make the figure). This figure highlights a number of areas that may be especially important in gesture processing, specifically with respect to processing the meanings they express.

The pars triangularis of the inferior frontal gyrus (IFGTr) is one area thought to be important for interpreting meaning communicated in language that might play a similar role in gesture processing. As it relates to language function, the IFGTr has been proposed to be involved in semantic retrieval and control processes when people interpret sentences and narratives (Thompson-Schill et al., 1997; Thompson-Schill, 2003). The IFGTr is consistently found to be active when people make overt semantic decisions in language tasks (Binder et al., 1997; Friederici et al., 2000; Devlin et al., 2003). A number of findings suggest this area also responds when interpreting a gesture's meaning.

The IFGTr may play a similar role in recognizing meaning from gestures as it does in verbal language. For example, bilateral IFGTr activity is stronger when people see co-speech gestures than when they process speech without gesture (Kircher et al.,

FIGURE 2 | Inferior frontal and lateral temporal regions relating gesture meaning. Abbreviations: MTGp, posterior middle temporal gyrus; STGp, posterior superior temporal gyrus; STGa, anterior superior temporal gyrus; IFGTr, pars triangularis; IFGOp, pars opercularis.

2009). Another study found greater left IFGTr activity when people observed incongruent speech and gestures than congruent (e.g., a speaker performed a "writing" gesture but said "hit"; Willems et al., 2007). However, under similar conditions in other studies – when hand movements that accompany speech are unrelated to the spoken content – activity has been found to be greater in the *right* IFGTr (Dick et al., 2009), or in both right and left IFGTr (Green et al., 2009; Straube et al., 2009). With emblems, not all studies report IFGTr activity. In those that do, though, it is found bilaterally (Lotze et al., 2006; Villarreal et al., 2008). Considered together, these results suggest that IFGTr may function similarly when people process gestures as it does when people process language. That is, IFGTr responses may be tuned to interpreting semantic information, particularly when it is necessary to unify a meaning expressed in multiple forms (e.g., via gesture and speech). When a gesture's meaning does not match speech, there is a strong IFGTr response. This may reflect added processing needed to reconcile a dominant meaning from mismatched speech and gesture. In contrast, when gesture and speech are congruent, understanding a message's meaning is more straightforward. This would rely less on regions that are particularly important for reconciling meaning from multiple representations.

Posterior to IFGTr, the pars opercularis (IFGOp) has also been found to respond when people process gestures. Anatomically positioned between IFGTr and PMv, IFGOp function has been associated with both language and motor processes. For example, this region is sensitive to audiovisual speech (Miller and D'Esposito, 2005; Hasson et al., 2007) and speech accompanied by gestures (Dick et al., 2009; Green et al., 2009; Kircher et al., 2009), as well as mouth and hand actions without any verbal communication (for review, see Binkofski and Buccino, 2004; Rizzolatti and Craighero, 2004). In other words, this area has a role in a number of language and motor functions. This includes comprehending verbal and motor information from both the mouth and the hands. Also, left frontoparietal lesions that involve the left IFG have been linked to impaired action recognition. Such impairment includes even when patients are asked to recognize an action via sounds

typically associated with the action (Pazzaglia et al., 2008a,b). Put simply, the IFGOp exhibits sensitivity in response to many types of information. Such broad sensitivity suggests the IFGOp as a site where integrative processes may be important in its function.

Yet, the inferior frontal cortex functions within a broader network. During language comprehension, this area interacts with lateral temporal cortex via the extreme capsule and uncinate fasciculus fiber pathways (Schmahmann et al., 2007; Petrides and Pandya, 2009), and potentially with posterior superior temporal cortex via the superior longitudinal fasciculus (Catani et al., 2005; Glasser and Rilling, 2008). fMRI studies have described strong functional connectivity between inferior frontal and lateral temporal areas in the human brain (Homae et al., 2003; Duffau et al., 2005; Mechelli et al., 2005; Skipper et al., 2007; Saur et al., 2008; Warren et al., 2009; Xiang et al., 2010). The lateral temporal cortex also responds stronger to speech with accompanying gestures than to speech alone. In particular, the posterior superior temporal sulcus (STSp) exhibits responses to visual motion, especially when it is biologically relevant (Bonda et al., 1996; Beauchamp et al., 2002). This also applies when people perceive gestures. But according to Holle et al. (2008), responses in STSp show sensitivity beyond just perceiving biological motion. They report that left STSp is more active when people see co-speech gestures than when they see speech with adaptor movements (such as adjusting the cuff of a shirt). In a subsequent study, Holle et al. (2010) report bilateral STSp activation when people see iconic co-speech gestures compared to when people see speech, gestures alone, or to audibly degraded speech. These authors posit the left STSp as a site where "integration of iconic gestures and speech takes place." However, their effects are not replicated in other studies (e.g., Willems et al., 2007, 2009; Dick et al., 2009). For example, Dick et al. (2009) found that bilateral STSp is active both for co-speech gestures and adaptor movements. Importantly, Dick et al. (2009) did not find that activity differed between co-speech gestures and adaptor movements. That is, they did not find evidence that the STSp is responsive to the semantic content of the hand movements. This is in line with the more recognized view that the STSp is generally responsive to biological motion.

In contrast to STSp, posterior middle temporal gyrus (MTGp) and anterior superior temporal cortex (STa) responses may be tuned to interpreting meaning, including when it is conveyed in gesture. For example, bilateral MTGp activity is stronger when people see metaphoric (Kircher et al., 2009) or iconic (Green et al., 2009; Willems et al., 2009) gestures than when they see either speech or gestures alone. In response to emblems, MTGp activity has been found in each the left (Lui et al., 2008; Villarreal et al., 2008) and right (Nakamura et al., 2004) hemispheres, as well as bilaterally (Lotze et al., 2006; Xu et al., 2009). Lesion studies have also corroborated this area's importance in recognizing an action's meaning (Kalenine et al., 2010). The MTGp was considered by some authors to be part of visual association cortex (von Bonin and Bailey, 1947; Mesulam, 1985). But this region's responses to auditory stimuli are also well documented (Zatorre et al., 1992; Wise et al., 2000; Humphries et al., 2006; Hickok and Poeppel, 2007; Gagnepain et al., 2008). Many studies have also associated MTGp activity with recognizing word meaning (Binder et al., 1997; Chao et al., 1999; Gold et al., 2006). Moreover, the semantic functions

of this region might not be modality dependent (e.g., related to verbal input). That is, the MTGp may have a role in interpreting meaning at a conceptual level. This view aligns with the results of a recent meta-analysis that characterizes the MTGp as "hetero-modal cortex involved in supramodal integration and conceptual retrieval" (Binder et al., 2009).

Many studies also implicate the STa in co-speech gesture processing (Skipper et al., 2007, 2009; Green et al., 2009; Straube et al., 2011). Activity in this region has been found for emblem processing, as well (Lotze et al., 2006). Changes in effective connectivity between STa and premotor cortex are found when people view gestures during story comprehension (Skipper et al., 2007). In language tasks, responses in this region have been associated with processing combinatorial meaning – usually as propositional phrases and sentences (Noppeney and Price, 2004; Humphries et al., 2006; Lau et al., 2008; Rogalsky and Hickok, 2009). A very similar function may be involved when people process gestures. After all, emblems convey propositional information that is easily translated to short spoken phrases. And co-speech gestures are typically processed in the context of sentence structures (Kircher et al., 2009) or full narratives (Skipper et al., 2007).

It appears that parts of the inferior frontal and lateral temporal cortices respond regardless of whether people perceive meaning represented verbally or manually. These areas' function suggests a shared neural basis for interpreting speech and gestures. This potentially shared basis is in line with the proposal that speech and gestures use a unified communication system (McNeill, 1992). When people must determine meaning among competing or ambiguous representations, anterior inferior frontal responses are most prevalent. In contrast, the MTGp responds strongly to represented meaning. In particular, this region appears to have a role at the level of conceptual recognition. The STa also functions in meaning recognition. Though it may be more important at the propositional level. That is, STa responses appear prominent when the expressed information involves units combined as a whole (e.g., as words are combined into phrases and sentences, or symbolic actions are associated with verbal complements).

DISTRIBUTED RESPONSES, DYNAMIC INTERACTIONS

Many reports in the gesture literature describe higher-level, complex functions (e.g., semantic integration) as localized to particular brain areas. However, the brain regularly exhibits responses that are highly distributed and specialized. These reflect the brain's dynamic functioning. Importantly, dynamic and distributed neural processes are facilitated by extensive functional connectivity and interactions. In this section, we first discuss how specialized distributed responses may apply in gesture processing, specifically in relation to motor system function. We then discuss the importance of understanding the functional relationships that facilitate cognitive processes, as these may be central in integrating and interpreting meaning from gestures and language.

The brain regularly exhibits widely distributed and diverse sets of responses. To more completely account for brain function in processing gestures, the meanings gestures convey, as well as communication in general, these distributed and diverse responses must be appreciated. Such responses may, in fact, comprise different levels of specialization that allow a functionally dynamic basis

for interpretation. One view suggests that meaning, at least as it pertains to action information, is encoded via corollary processes between action and language systems (Pulvermuller, 2005; Pulvermuller et al., 2005). This view postulates that a correlation between action and action-related language leads to functional links between them. These links result in this information's encoding by distributed and interactive neural ensembles. The often cited example used to support this view is that processing effector-specific words (e.g., kick, lick, pick) involves brain activity in areas used to produce the effector-specific actions (e.g., with the leg, tongue, and mouth, respectively; Hauk et al., 2004). As reviewed above, gesture processing does, in fact, incorporate motor area responses. Whether particular types of semantic meaning conveyed by gestures is represented via distinct, distributed neural ensembles – similar to what is found for words that represent effector-specific information – is uncertain. Conceivably, motor responses in gesture processing could reflect the brain's sensitivity to represented features, beyond a gesture's visuomotor properties. In other words, gestures that symbolically represent motor information (e.g., a gesture used to represent a specific body part, such as the leg) could also rely on a somatotopic encoding analogous to that found for words that represent body parts. The way that the brain would achieve this degree of specialization is uncertain. It is increasingly clear though that responses to gestures are, indeed, diverse and distributed among distinct regions (**Figures 1** and **2**). Yet, this view that action information is encoded via corollary processes between action and language systems does not account for processing meaning that does not involve action (e.g., "The capitol is Sacramento"). Thus, while distributed encoding in the motor system may play a part in processing information that relates actions to the effectors used to perform actions, accounting for how the brain interacts with meaning more generally requires a broader basis.

Understanding how the brain interprets gestures and the information they convey requires appreciating the way that the brain represents information and implements higher-level functions. This requires characterizing not just the function of distinct locations that may show tuning to particular features, but also the dynamic interactions and aggregate function of distributed responses (McIntosh, 2000). Certain brain areas (e.g., in sensory and motor cortices) may be specialized to respond to particular kinds of information. But higher-level processes, such as memory and language (and by extension, interpreting meaning), require understanding the way that the brain relates and integrates information. Most of the previously discussed studies, particularly those focused on co-speech gestures, have aimed to characterize the neural integration of gesture and language processing. Many localize this process with results for sets of individual regions. Some of the implicated regions include the IFG (Willems et al., 2007; Straube et al., 2009), temporo-occipital junction (Green et al., 2009), and STSp (Holle et al., 2008, 2010). However, to characterize complex, integrative functions by one-to-one alliances with individual regions, without also acknowledging those neural mechanisms that might enable relationships among particular regions, loses sight of the brain's dynamic and interconnected nature. For example, the same brain areas may exhibit activity in different tasks or in response to similar information from different

mediums (e.g., each symbolic gestures and spoken language; Xu et al., 2009). Similarly, the brain can exhibit distributed function that is evoked by presentation from the same medium (e.g., co-speech gestures). Thus, whereas a particular brain area may be similarly active across what seem to be different cognitive tasks, what "distinguishes [these] tasks is the pattern of spatiotemporal activity and interactivity more than the participation of any particular region" (McIntosh, 2000). Some previous gesture work has examined the functional relationships among anatomically diverse areas' responses (e.g., Skipper et al., 2009; Willems et al., 2009; Xu et al., 2009). However, such analyses are still the exception in the gesture literature. A more global perspective that recognizes specialized responses interact across the whole brain to implement cognitive processes is needed. That is, whereas one neural system might be particularly tuned to process gestures, another might be better tuned to process verbal discourse. Importantly, while such systems may organize with varying degrees of specialization, it is their dense interconnectivity that enables dynamic neural processing in context. This may be especially important for understanding the way that the brain functions in the perceptually rich and complex scenarios that comprise typical experience. Thus, to understand the way the brain implements complex processes, such as integrating and interpreting meaning from gestures and language, "considering activity of the entire brain rather than individual regions" (McIntosh, 2000) is vital.

RELEVANCE FOR REAL WORLD INTERACTIONS (BEYOND THE EXPERIMENT)

To understand a gesture, an observer must relate multiple pieces of information. For example, emblem comprehension involves visually perceiving the gesture, as well as processing its meaning. Similarly, interpreting co-speech gestures requires visual perception of the gesture. But, in contrast with emblems, co-speech gesture processing involves associating the action with accompanying auditory verbal information. Importantly, speech and gesture information do not combine in an additive way. Rather, these sources *interactively* contribute meaning, as people integrate them into a unified message (Kelly et al., 1999; Bernardis and Gentilucci, 2006; Gentilucci et al., 2006). There is also pragmatic information that is part of the natural context in which these actions are typically experienced. This pragmatic context can also influence a gesture's interpretation (Kelly et al., 1999, 2007). Another factor that can impact interpretation is the observer's intent. For example, brain responses to the same gestures can differ depending on whether an observer's goal is to recognize the hand as meaningful or, categorically, as simply a hand (Nakamura et al., 2004). Therefore, to comprehensively appreciate the way that the brain processes gestures – particularly, the meaning they express – these diverse information sources should be accounted for in ways that recognize contextual influences.

However, most fMRI studies of gesture processing present participants with stimuli that have little or no resemblance to anything they would encounter outside of the experiment. Of course, researchers do this with the intention of isolating brain responses to a specific feature or function of interest by controlling for all other factors. Some examples in prior gesture studies include having the person performing the gestures cloaked in all black (Holle

et al., 2008), allowing only the actor's hand to be visible through a screen (Montgomery et al., 2007), and putting a large circle that changes colors on the actor's chest (Kircher et al., 2009). Such unusual visual information could pose a number of problems. Most concerning is that it could distract attention from the visual information that is relevant and of interest (e.g., the gestures). The inverse is also possible, however. That is, irregular visual materials might artificially enhance attention toward the gestures. In either case, such materials do not generalize to people's typical experience.

Beyond the materials' visual aspects, many experiments have also used conditions that are explicitly removed from familiar experience. For example, a common approach has been to compare responses to co-speech gestures with responses to speech and gestures that do not match. The idea here is that the difference between these conditions would reveal brain areas involved in "integration" (itself often only loosely or not at all defined). The reasoning is that in the condition where gesture and speech are mismatched the brain is presumed to respond to each as dissociated signals. But, when gesture and speech are congruent, there is recognition of a unified representation or message. However, meaningless hand actions evoke a categorically different brain response than meaningful ones (Decety et al., 1997). Thus, interpreting these findings can be difficult. Such an approach also brings to light an additional potential limitation: Many results are determined by simply subtracting responses collected in one condition from those in another condition. In other words, results are often achieved by subtracting responses generated under exposure to one input (e.g., speech) from responses to a combination of inputs (e.g., speech and gesture). The difference in activity for the contrast or condition of interest is then typically described as the effect. Not only does such an approach make it difficult to characterize the interaction between speech and gesture that gives a co-speech gesture meaning, but it also assumes each the brain and fMRI signals are linear (which they are not; Logothetis et al., 2001). Thus, results derived under such conditions can be hard to interpret, particularly as to the degree to which they inform gesture processing. They may also yield results that are hard to replicate, even when a study explicitly tries to do so (Dick et al., 2009).

Another issue is that many researchers require their participants to do motor tasks (such as pushing buttons to record behaviors) that are accessory to the function of interest during fMRI data collection. A number of prior gesture studies have used these tasks (e.g., Green et al., 2009; Kircher et al., 2009). However, having participants engage in motor behaviors while in the scanner could potentiate responses in a confounding way. In other words, motor responses could then be due to the motor behavior in the accessory task, as well as interfere with potential motor responses relevant to processing the gestures (the "motor output problem"; Small and Nusbaum, 2004). This can be especially problematic when motor areas are of primary interest. Thus, these accessory tasks can generate brain responses that are hard to disentangle from those that the researchers intended to examine.

To avoid many of these potential limitations, more naturalistic conditions need to be considered in studying gesture and language function. One immediate concern may be that evaluating data collected under contextualized, more naturalistic exposures can pose

a challenge for commonly used fMRI analysis measures. The most typical approach for analyzing fMRI data uses the general linear model. This requires an *a priori* specified hemodynamic response model against which the collected responses can be regressed. Also, particularly for event-related designs, an optimized stimulus event sequence is needed to avoid co-linearity effects that may mask signal of interest from co-varying noise-related artifacts. With naturalistic, continuously unfolding stimuli, meeting such requirements is not always possible. Fortunately, many previous authors have demonstrated approaches that achieve systematic, informative results from data collected under more naturalistic conditions (e.g., Zacks et al., 2001; Bartels and Zeki, 2004; Hasson et al., 2004; Mathiak and Weber, 2006; Malinen et al., 2007; Spiers and Maguire, 2007; Yarkoni et al., 2008; Skipper et al., 2009; Stephens et al., 2010). The intersubject synchronization approach used to analyze data collected while people watched segments of "The Good, The Bad, and The Ugly" is probably the most well-known example (Hasson et al., 2004). However, others researchers have successfully derived informative fMRI results from data collected as people comprehended naturalistic audiovisual stories (Wilson et al., 2008; Skipper et al., 2009), read narratives (Yarkoni et al., 2008), watched videos that presented everyday events such as doing the dishes (Zacks et al., 2001), and had verbal communication in the scanner (Stephens et al., 2010). These achievements are important because they demonstrate ways to gainfully use context rather than unnaturally remove it. Considering the interactive, integrative, and contextual nature of gesture and language processing – particularly in typical experience – it is essential to consider such approaches as the study of gesture and language moves forward.

SUMMARY AND FUTURE PERSPECTIVES
SUMMARY
Current findings indicate that two types of brain areas are implicated when people process gestures, particularly emblems and co-speech gestures. One set of areas comprises parietal and premotor regions that are important in processing hand actions. These areas are sensitive to both gestures and actions that are not directly symbolic, such as grasping an object. Thus, it is likely that the function of these parietal and premotor regions primarily involves perceiving hand actions, rather than interpreting their meaning. In contrast, the other set of areas includes inferior frontal and lateral temporal regions. These regions are classically associated with language processing. They may function in a similar capacity to process symbolic meaning conveyed with gestures. While the current data present a general consensus for these areas' roles, the way that the brain reconciles manual and symbolic information it is not yet clear.

The lack of reconciliation between the neurobiology of action understanding and that of understanding symbolic information yields at least two prominent points concerning gesture research. First, a characteristic response profile for gesture processing remains unspecified. That is, among results for these two sets of areas, there is a strong degree of variability. This variability clouds whether certain regions' responses are central in gesture processing. It also obscures whether there are particular sets of responses that implement the integrative and interpretive

mechanisms needed to comprehend gestures and language. Second, there is currently minimal exposition at the neural systems level, particularly that relates the brain's anatomical and functional interconnections. The variable and widely distributed responses found in previous gesture studies suggest a broader neural perspective is needed. In other words, function throughout the brain and its interconnectivity must be considered. In the final section, we discuss important issues for moving forward in the study of gesture and language processing and then relate them to some outstanding topics.

CONTEXT IS PERVASIVE
To move forward in understanding the way that the brain processes gestures and language, the fundamental importance of context must be recognized. This pertains both to experimental design and as a principle of brain function ("neural context"; McIntosh, 2000). Here, we will first briefly summarize the importance of considering context as it relates to experimental approaches. We then follow with a discussion of context as it relates to understanding interactive and dynamic function across the whole brain.

Concerning the role of context in experimental design, one of the primary hurdles in using contextualized, naturalistic materials is that they do not typically satisfy the *a priori* requirements of many commonly used imaging analysis methods (as discussed above). However, systematically evaluating fMRI data collected under contextualized, continuous exposures is achievable. Prior imaging work includes numerous informative results derived from fMRI data collected under more naturalistic conditions (Zacks et al., 2001; Bartels and Zeki, 2004; Hasson et al., 2004; Mathiak and Weber, 2006; Malinen et al., 2007; Spiers and Maguire, 2007; Yarkoni et al., 2008; Skipper et al., 2009; Stephens et al., 2010). Gesture and language researchers should increasingly consider such methods. Applying these methods could supplement subtractive approaches (both in designing experimental conditions and their analysis) that might mischaracterize the way the brain operates. In particular, these methods might provide better insight into integrative mechanisms in gesture and language processing, as they are implemented in typical experience.

Concerning the role of context as it relates to brain function, it is important to maintain perspective of the entire brain's function. Gesture and language processing exemplifies this need, as they incorporate responses that are not only diverse and distributed across the brain but are also interactive and interconnected. In other words, recognizing that the brain is a complex system in the formal (mathematical) sense will benefit efforts to understand gesture and language function. Accordingly, this will necessitate investigations to focus on distributed neural systems, rather than just on localizing complex processes to individual regions. Examining the neural relationships between regions could more comprehensively characterize gesture and language processing. In any case, it would promote research that better investigates those neural properties that facilitate higher-level functions. For example, the distributed brain networks involved in functions such as language, memory, and attention comprise multiple pathways (Mesulam, 1990) with semi-redundant and reciprocal connectivity (Tononi and Sporns, 2003; Friston, 2005). These connections may involve regions with varying degrees

of specialization to particular information types (e.g., gesture and/or language stimuli). Also, a particular region's specialization may be determined, in part, by its connectivity (McIntosh, 2000). Put simply, more than one region may be engaged in a particular function, and more than one function may engage a particular region. It is important then to consider that "specialization is only meaningful in the context of functional integration and vice versa" (Friston, 2005). Therefore, to appreciate how the brain implements complex operations (e.g., information integration, interpretation), insight into neural function at multiple levels of representation is needed. This implicates not just the level of regional specialization (the system's elements) but also the anatomical and functional relationships that facilitate these elements' interactive and distributed processes (their connections).

FUNCTIONALLY INTERACTIVE AND DISTRIBUTED OPERATIONS

Recognizing the importance of context – both as it relates to experimental design and as a basic principle of brain function – will allow future studies to better examine how the brain operates through interactive and distributed function. This will encourage researchers to ask questions about the brain that incorporate contextual factors, rather than artificially eliminate them. This is important because people function in a world that requires continual interaction with abundant, changing, and diverse information sources. A greater degree of resemblance and relevance to the real world can and should be incorporated into future experimental designs. Below, we consider a few outstanding issues in the study of gesture and language processing for which these approaches may be especially useful.

One issue is whether the neural mechanisms involved in perceiving gestures distinguish among the diverse semantic meanings they can represent. For example, co-speech gestures can be used to represent different types of information such as physical objects, body parts, or abstracts ideas. It is unclear if the brain exhibits specialized responses that are particular to these meanings. Whether or not such responses might be distributed in distinct regions is also uncertain. Additionally, it is unclear how distributed responses to gestures would interact with other neural systems to incorporate contextual factors.

Another issue is to what extent pragmatics and situational factors influence how brain systems organize in processing gesture and language information. Pragmatic knowledge does, in fact, play a role in gesture comprehension (Kelly et al., 1999, 2007). It is uncertain, however, to what degree certain systems, such as the putative action understanding circuit relating parietal and premotor responses, would maintain a functional role in perceiving gestures under varying situational influences. For example, a person might perceive a pointing gesture by someone yelling "Over there!" while trying to escape a burning building. This is quite different, and would probably involve different neural systems, than perceiving the same gesture and language conveyed to indicate where the TV remote is located. Similarly, the way that neural mechanisms coordinate as a function of the immediate verbal context is also uncertain. Recall the example provided in the Introduction of this paper: A person moving their hand in a rolling motion can represent one thing accompanied by "The wheels are turning" but another when accompanied by "The meeting went on and on." To process such information, neural mechanisms that enable functional interaction among responses to the gesture and accompanying verbal content, and that dynamically implement interpreting their meaning in context, would need to be incorporated. Such interactive neural mechanisms are currently unclear and deserve further investigation. Thus, addressing these issues would further inform the neural basis of gesture processing, as well as how the brain might encode and interpret meaning.

In conclusion, the current gesture data implicate a number of brain areas that to differing extents index action and symbolic information processing. However, the neural relationships that provide a dynamic and interactive basis for comprehension need to be accounted for, as well. This will allow a more complete look at the way that the brain processes gestures and their meanings. As gesture and language research moves forward, a vital factor that needs further consideration is the role of context. The importance of context applies both to the settings in which people process gesture and language information, as well as to understanding in what way distributed and interconnected responses throughout the brain facilitate this information's interactive comprehension.

REFERENCES

Bartels, A., and Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Hum. Brain Mapp.* 21, 75–85.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159.

Bernardis, P., and Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia* 44, 178–190.

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19, 2767–2796.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362.

Binkofski, F., and Buccino, G. (2004). Motor functions of the Broca's region. *Brain Lang.* 89, 362–369.

Bonda, E., Petrides, M., Ostry, D., and Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J. Neurosci.* 16, 3737–3744.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., and Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404.

Buccino, G., Binkofski, F., and Riggio, L. (2004). The mirror neuron system and action recognition. *Brain Lang.* 89, 370–376.

Cassell, J., McNeill, D., and McCullough, K.-E. (1999). Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics Cogn.* 7, 1–34.

Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann. Neurol.* 57, 8–16.

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

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., and Fazio, F. (1997). Brain activity during observation of actions – influence of action content and subject's strategy. *Brain* 120, 1763–1777.

Devlin, J. T., Matthews, P. M., and Rushworth, M. F. S. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *J. Cogn. Neurosci.* 15, 71–84.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events – a neurophysiological study. *Exp. Brain Res.* 91, 176–180.

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., and Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum. Brain Mapp.* 30, 3509–3526.

Duffau, H., Gatignol, P., Mandonnet, E., Peruzzi, P., Tzourio-Mazoyer, N., and Capelle, L. (2005). New insights into the anatomo-functional connectivity of the semantic system: a study using cortico-subcortical electrostimulations. *Brain* 128, 797–810.

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal communication: categories, origins, usage, and coding. *Semiotica* 1, 49–98.

Fabbri-Destro, M., and Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology (Bethesda)* 23, 171–179.

Feyereisen, P. (2006). Further investigation on the mnemonic effect of gestures: their meaning matters. *Eur. J. Cogn. Psychol.* 18, 185–205.

Fogassi, L., Gallese, V., Fadiga, L., and Rizzolatti, G. (1998). Neurons responding to the sight of goal directed hand/arm actions in the parietal area PF (7b) of the macaque monkey. *Soc. Neurosci. Abstr.* 24, 257.5.

Friederici, A. D., Opitz, B., and Cramon, D. Y. V. (2000). Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. *Cereb. Cortex* 10, 698–705.

Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836.

Gagnepain, P., Chetelat, G., Landeau, B., Dayan, J., Eustache, F., and Lebreton, K. (2008). Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. *J. Neurosci.* 28, 5281–5289.

Gentilucci, M., Bernardis, P., Crisi, G., and Dalla Volta, R. (2006). Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J. Cogn. Neurosci.* 18, 1059–1074.

Geyer, S., Matelli, M., Luppino, G., and Zilles, K. (2000). Functional neuroanatomy of the primate isocortical motor system. *Anat. Embryol.* 202, 443–474.

Glasser, M. F., and Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cereb. Cortex* 18, 2471–2482.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., and Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *J. Neurosci.* 26, 6523–6532.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends Cogn. Sci. (Regul. Ed.)* 3, 419–429.

Goldin-Meadow, S. (2003). *Hearing Gesture: How our Hands Help us Think*. Cambridge, MA: Belknap Press of Harvard University Press.

Goldin-Meadow, S. (2006). Talking and thinking with our hands. *Curr. Dir. Psychol. Sci.* 15, 34–39.

Grafton, S. T., Arbib, M. A., Fadiga, L., and Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp. Brain Res.* 112, 103–111.

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., and Kircher, T. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30, 3309–3324.

Grezes, J., Armony, J. L., Rowe, J., and Passingham, R. E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage* 18, 928–937.

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

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640.

Hasson, U., Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2007). Abstract coding of audiovisual speech: beyond sensory representation. *Neuron* 56, 1116–1126.

Hauk, O., Johnsrude, I., and Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024.

Holle, H., Obleser, J., Rueschemeyer, S.-A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49, 875–884.

Homae, F., Yahata, N., and Sakai, K. L. (2003). Selective enhancement of functional connectivity in the left prefrontal cortex during sentence processing. *Neuroimage* 20, 578–586.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychol. Bull.* 137, 297–315.

Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037.

Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.* 18, 665–679.

Kalenine, S., Buxbaum, L. J., and Coslett, H. B. (2010). Critical brain regions for action recognition: lesion symptom mapping in left hemisphere stroke. *Brain* 133, 3269–3280.

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

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

Kelly, S. D., Ward, S., Creigh, P., and Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain Lang.* 101, 222–233.

Kendon, A. (1994). Do gestures communicate? A review. *Res. Lang. Soc. Interact.* 27, 175–200.

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., and Green, A. (2009). Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. *Neuropsychologia* 47, 169–179.

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933.

Leiguarda, R. C., and Marsden, C. D. (2000). Limb apraxias: higher-order disorders of sensorimotor integration. *Brain* 123(Pt 5), 860–879.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.

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

Lui, F., Buccino, G., Duzzi, D., Benuzzi, F., Crisi, G., Baraldi, P., Nichelli, P., Porro, C. A., and Rizzolatti, G. (2008). Neural substrates for observing and imagining non-object-directed actions. *Soc. Neurosci.* 3, 261–275.

Malinen, S., Hlushchuk, Y., and Hari, R. (2007). Towards natural stimulation in fMRI – issues of data analysis. *Neuroimage* 35, 131–139.

Manthey, S., Schubotz, R. I., and Von Cramon, D. Y. (2003). Premotor cortex in observing erroneous action: an fMRI study. *Brain Res. Cogn. Brain Res.* 15, 296–307.

Mathiak, K., and Weber, R. (2006). Toward brain correlates of natural behavior: fMRI during violent video games. *Hum. Brain Mapp.* 27, 948–956.

McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Netw.* 13, 861–870.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

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

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. A., Patterson, K., McClelland, J. L., and Price, C. J. (2005). Dissociating reading processes on the basis of neuronal interactions. *J. Cogn. Neurosci.* 17, 1753–1765.

Mesulam, M. M. (1985). "Patterns in behavioral neuroanatomy: association areas, the limbic system, and hemispheric specialization," in

Principles of Behavioral Neurology, ed. M. M. Mesulam (Philadelphia: F. A. Davis), 1–70.

Mesulam, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* 28, 597–613.

Miller, L. M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893.

Montgomery, K. J., Isenberg, N., and Haxby, J. V. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Soc. Cogn. Affect. Neurosci.* 2, 114–122.

Nakamura, A., Maess, B., Knosche, T. R., Gunter, T. C., Bach, P., and Friederici, A. D. (2004). Cooperation of different neuronal systems during hand sign recognition. *Neuroimage* 23, 25–34.

Noppeney, U., and Price, C. J. (2004). An FMRI study of syntactic adaptation. *J. Cogn. Neurosci.* 16, 702–713.

Pazzaglia, M., Pizzamiglio, L., Pes, E., and Aglioti, S. M. (2008a). The sound of actions in apraxia. *Curr. Biol.* 18, 1766–1772.

Pazzaglia, M., Smania, N., Corato, E., and Aglioti, S. M. (2008b). Neural underpinnings of gesture discrimination in patients with limb apraxia. *J. Neurosci.* 28, 3030–3041.

Perani, D., Fazio, F., Borghese, N. A., Tettamanti, M., Ferrari, S., Decety, J., and Gilardi, M. C. (2001). Different brain correlates for watching real and virtual hand actions. *Neuroimage* 14, 749–758.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi:10.1371/journal.pbio.1000170

Pierno, A. C., Becchio, C., Wall, M. B., Smith, A. T., Turella, L., and Castiello, U. (2006). When gaze turns into grasp. *J. Cogn. Neurosci.* 18, 2130–2137.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.

Pulvermuller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rizzolatti, G., and Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. *Curr. Opin. Neurobiol.* 18, 179–184.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Curr. Opin. Neurobiol.* 12, 149–154.

Rogalsky, C., and Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cereb. Cortex* 19, 786–796.

Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.

Shmuelof, L., and Zohary, E. (2005). Dissociation between ventral and dorsal fMRI activation during object and action recognition. *Neuron* 47, 457–470.

Shmuelof, L., and Zohary, E. (2006). A mirror representation of others' actions in the human anterior parietal cortex. *J. Neurosci.* 26, 9736–9742.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Curr. Biol.* 19, 1–7.

Small, S. L., and Nusbaum, H. C. (2004). On the neurobiological investigation of language understanding in context. *Brain Lang.* 89, 300–311.

Spiers, H. J., and Maguire, E. A. (2007). Decoding human brain activity during real-world experiences. *Trends Cogn. Sci. (Regul. Ed.)* 11, 356–365.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14425–14430.

Straube, B., Green, A., Bromberger, B., and Kircher, T. (2011). The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Brain Mapp.* 32, 520–533.

Straube, B., Green, A., Weis, S., Chatterjee, A., and Kircher, T. (2009). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cogn. Neurosci.* 21, 821–836.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia* 41, 280–292.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797.

Tononi, G., and Sporns, O. (2003). Measuring information integration. *BMC Neurosci.* 4, 31. doi:10.1186/1471-2202-4-31

Tremblay, P., and Small, S. L. (2011). From language comprehension to action understanding and back again. *Cereb. Cortex* 21, 1166–1177.

Villarreal, M., Fridman, E. A., Amengual, A., Falasco, G., Gerscovich, E. R., Ulloa, E. R., and Leiguarda, R. C. (2008). The neural substrate of gesture recognition. *Neuropsychologia* 46, 2371–2382.

von Bonin, G., and Bailey, P. (1947). *The Neocortex of Macaca Mulatta.* Urbana: University of Illinois Press.

Warren, J. E., Crinion, J. T., Lambon Ralph, M. A., and Wise, R. J. (2009). Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. *Brain* 132, 3428–3442.

Willems, R. M., Ozyurek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333.

Willems, R. M., Ozyurek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47, 1992–2004.

Wilson, S. M., Molnar-Szakacs, I., and Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex* 18, 230–242.

Wise, R. J., Howard, D., Mummery, C. J., Fletcher, P., Leff, A., Buchel, C., and Scott, S. K. (2000). Noun imageability and the temporal lobes. *Neuropsychologia* 38, 985–994.

Wu, Y. C., and Coulson, S. (2005). Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42, 654–667.

Xiang, H. D., Fonteijn, H. M., Norris, D. G., and Hagoort, P. (2010). Topographical functional connectivity pattern in the Perisylvian language networks. *Cereb. Cortex* 20, 549–560.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., and Braun, A. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20664–20669.

Yarkoni, T., Speer, N. K., and Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *Neuroimage* 41, 1408–1425.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., and Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.* 4, 651–655.

Zatorre, R. J., Evans, A. C., Meyer, E., and Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846–849.

APPENDIX
STUDIES USED IN FIGURES 1 AND 2
Emblematic gesture
1. Lotze et al. (2006)
2. Lui et al. (2008)
3. Montgomery et al. (2007)
4. Nakamura et al. (2004)
5. Villarreal et al. (2008)

Co-speech gesture
1. Dick et al. (2009)
2. Green et al. (2009)
3. Holle et al. (2008)
4. Holle et al. (2010)
5. Hubbard et al. (2009)
6. Kircher et al. (2009)
7. Skipper et al. (2007)
8. Skipper et al. (2009)
9. Straube et al. (2009)
10. Straube et al. (2011)
11. Willems et al. (2007)
12. Willems et al. (2009)

Unrelated co-speech gesture
1. Dick et al. (2009)
2. Green et al. (2009)
3. Holle et al. (2008)
4. Hubbard et al. (2009)
5. Straube et al. (2009)
6. Willems et al. (2007)
7. Willems et al. (2009)

Grasping
1. Buccino et al. (2001)
2. Buccino et al. (2004)
3. Grezes et al. (2003)
4. Grafton et al. (1996)
5. Shmuelof and Zohary (2005)
6. Shmuelof and Zohary (2006)
7. Manthey et al. (2003)
8. Perani et al. (2001)
9. Pierno et al. (2006)
10. Tremblay and Small (2011)