

Gesture use in social interaction: how speakers' gestures can reflect *listeners'* thinking

Holler, Judith & Beattie, Geoffrey
University of Manchester

judith.holler@manchester.ac.uk
geoff.beattie@manchester.ac.uk

Abstract

The question as to why we move our hands and arms while we speak has intrigued many researchers in the past, and it still does. However, there has been much debate concerning the cause and function of these spontaneous movements which often represent meaningful information. Some argue that imagistic gestures benefit mainly the speaker, while others argue that they predominantly serve to assist the communication of information to an interlocutor. Two experimental studies are presented in this paper, which examine the influence of social-interactive processes on iconic gestures. The first focuses on the use of gesture in association with speakers' clarification of verbal (lexical) ambiguity. The second study investigates the influence of common ground on gesture use. The findings obtained from these studies support the notion that social context does influence gesture and that speakers use iconic gestures for their interlocutors, i.e. because they *intend* to communicate.

Key-words: *Iconic gestures, gesture production, social interaction, ambiguity, common ground*

1. Introduction

When people talk they usually move their hands and arms while they speak. Many of these gestures are imagistic gestures. McNeill (1985) was amongst the first to point out that these images spontaneously created by our hands reveal important insights into speakers' thoughts. This is because, he argues, gesture and speech are tightly connected; they share an early computational stage in the process of utterance formation and the two sides remain in constant dialogue throughout this process. Imagery and linguistic content unfold together in what McNeill (e.g. 1992, 2005) refers to as a dialectic process. The end product is an utterance that comprises a linguistic side expressed in speech, as well as an imagistic side expressed in gesture. Therefore, the verbal components of the utterances speakers produce contain only part of the message a speaker is trying to convey, and the imagistic hand gestures accompanying these verbal components can add considerable amounts of semantic information to the speech (e.g. McNeill 1992; for experimental evidence for the communicative effects of gestures in addition to speech see Beattie, 2003 and Beattie & Shovelton 1999a, 1999b, 2001, 2002).

Although research has shown that imagistic hand gestures can communicate, why speakers make these gestures is still a much debated issue. Some researchers argue that communicative effects of gestures are merely accidental and not intended (e.g. Butterworth & Hadar, 1989; Krauss, Chen, & Gottesman, 2000; Krauss, Morrel-Samuels, & Colasante, 1991) and that, instead, the main function of these gestures is to facilitate lexical retrieval and thus to benefit

the speaker rather than the listener. Other researchers have argued against this, opining that speech-accompanying hand gestures *are* communicatively intended and strongly influenced by conversational context (e.g. Bavelas & Chovil, 2000; Kendon, 1983, 1985, 2004).

Several investigations have provided experimental evidence that suggests that speakers do indeed produce gestures for their addressees. For example, Beattie & Aboudan (1994) found that speakers produce more imagistic gestures in dialogic interaction than when they talk in monologue. Bavelas, Kenwood, Johnson & Phillips (2002) showed that speakers produce more imagistic hand gestures when they were told that a video made of them while describing certain stimuli would be shown to other people than when they were told that an audio recording would be played to the other participants. Furthermore, the gestures they produced in the latter condition were more redundant with the speech (i.e. they added less information) than those produced in the former. Furuyama (2000) examined hand gestures made by teachers and learners in an origami task which, amongst others, revealed that speakers specifically oriented certain gestures that they used in this context towards their addressee. Further evidence comes from an investigation by Özyürek (2000, 2002); in these studies, she analysed speakers' use of shared gesture space when talking to one or two addressees, and when talking to addressees that were either located opposite or towards the side of the speaker. The analysis showed that speakers alter the way they represent certain motion events in gesture space by taking into account how their own and their interlocutors' gesture space, constituting part of the social-interactional context, intersect.

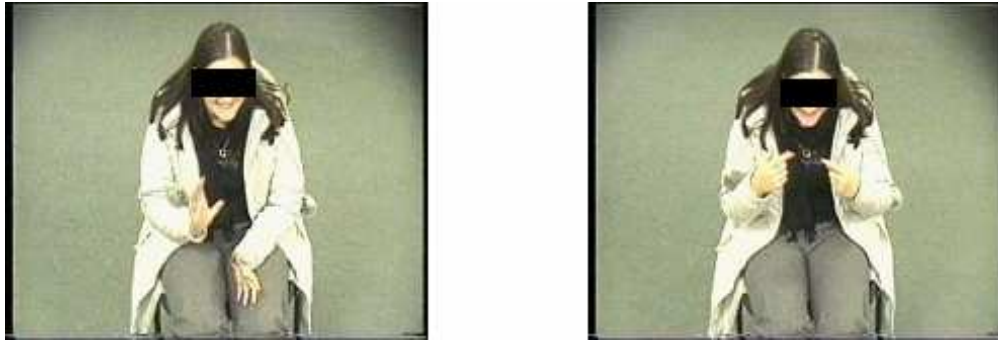
These studies provide important first insights into the effects social-interactional processes have on gesture use and to what extent speakers produce gestures for their addressees. What the research does show is that both the presence of an addressee and dialogue between interactants affect the frequency of gestures, and the former also affects the way in which gesture and speech interact in the representation of semantic information (in terms of the degree of redundancy or complementarity). It also shows that it affects how speakers represent information in terms of the form of gestures, their orientation and movement in gesture space. However, apart from physical co-presence (or visibility), spatial arrangements and the extent of verbal interactivity, an important question is to what extent speakers take into consideration their addressees' *thinking* when gesturing.

The two studies described in this paper examine this question. Study 1 uses lexical ambiguity as a test case to investigate whether speakers anticipate addressees' understanding problems and use gesture to provide semantic information to prevent these problems from occurring. Study 2 has a wider focus as it investigates the effect of 'common ground' (the knowledge that interactants in conversation share, e.g. Clark, 1996) on gesture use, i.e. the speaker's more general anticipation of the addressee's knowledge and thinking.

2. Study 1

In this experiment, 10 speakers were asked to reproduce sentences which contained homonyms and were globally ambiguous (e.g., 'The old man's glasses were filthy' [homonym: glasses; alternative interpretations: drinking glasses and spectacles]). They were then asked what the ambiguous sentence could mean in one sense and what in the other, which was intended to simulate a request for clarification often posed by addressees in everyday talk (such as, 'what do you mean?', or, 'do you mean x or y?').

Fig. 1: Participant using disambiguating gestures referring to the concept of ‘drinking glasses’ (left) and ‘spectacles’ (right).



The analysis focused on how speakers would deal with the ambiguities and how they would draw upon the two modalities, gesture and speech, in order to resolve them. The results showed that in 140 instances speakers recognised and attempted to resolve the ambiguity (either using only gesture, only speech or both). In 65 out of these 140 cases (46%), speakers used gesture to disambiguate what they were saying (in addition to or in absence of speech); regarding seven out of these 65 cases (11%; or 5% if the total amount of disambiguation attempts is considered) gesture was the only source of disambiguating information (i.e. the speech remained ambiguous, such as ‘it could mean glasses or it could mean glasses’, accompanied by two gestures, one with each mention of the word ‘glasses’, representing the concepts ‘drinking glasses’ and ‘spectacles’ (in many cases only one of the two meanings was not disambiguated verbally but only gesturally; two different meanings and their disambiguation were always counted as separate instances). In the remaining 133 cases (95%), speech was used in a disambiguating manner, and 58 of these 133 cases were accompanied by disambiguating hand gestures (44%). Thus, it appears that speech was used to disambiguate in the large majority of cases but gesture was used to disambiguate in addition to speech almost half of the time, and in some cases indeed as the only source of disambiguating information.¹

However, we know from past research that the very nature of dialogue can increase the frequency with which gestures are used by speakers (Beattie & Aboudan, 1994). Therefore, it could be that the requests for clarification placed by the addressee themselves encouraged the frequent gesture use. In order to test this, some of the homonyms were inserted into four different picture stories (created in a way that allowed for incorporating both alternative meanings of a homonym into the context of the story in close proximity), along with non-ambiguous control words. When asked to narrate the picture stories to interlocutors who did not know the story content, it was found that the ambiguous words were accompanied by a proportionally larger number of gestures, and this difference was statistically significant ($T=5, N=10, p<.02$).¹

This is quite clear evidence that speakers do anticipate their addressees’ thought process, at least when it comes to individual words that might cause confusion. However, does this mean that speakers take into account the wider conversational context when anticipating addressees’ thinking? Two individual examples we have come across seem to suggest that they do. The first one stems from the same study just described, more precisely in association with explaining the alternative meanings of the word ‘pot’. Whereas three of the participants

who gestured while explaining this particular ambiguity used their hands to represent the round, bowl-like shape of a cooking pot when contrasting it to the concept of marijuana, participant 7 (see Table 1) did something else. Instead of representing the pot as a container of some sort with a round element to it, this speaker imitated to be gripping an oblong-shaped handle with one hand. Such a handle is quite typical for English cooking pots (more so than one on either side of the pot). This variation may of course simply illustrate the idiosyncrasy that characterises imagistic gestural representations. However, another possibility is that the reason lies in the comparisons the speakers were making. Speakers 4, 8 and 10 compare the concept of a pot that represents a container (such as a pan/cooking pot, jug or plant pot) which is typically round and bowl shaped to a concept that shares neither of these qualities (i.e. the drug). In these cases, the gestures showing the round, bowl-like shape of the container are clearly disambiguating. However, speaker 7 considers *three* alternative interpretations, rather than just two. First, she refers to a flower pot, without an accompanying gesture. The concept of a flower pot usually is round and bowl-shaped in some sense. Then she refers to the concept of a cooking pot, or pan, and this reference is accompanied by a gesture. However, in order for this gesture to be disambiguating, it must represent something other than the round bowl-shape of the pan, since these features are shared with the concept of a flower pot. At this point the speaker uses a gesture which does exactly this – it represents the handle of a saucepan, a feature that is clearly not associated with either a flower pot or marijuana and thus is disambiguating.

If this variation in gesture is indeed the consequence of the speaker being aware as to what the semantic aspects of the individual concepts are which would be most effective in terms of disambiguation, this would suggest that when ‘designing’ their gestures, speakers take into account their addressees’ understanding and potential understanding problems. In this case, the speaker had to bear in mind that the addressee will have been thinking of a flower pot, and consider what the most effective gestural representation might be for differentiating this kind of pot from the concept of a cooking pot.

Table 1: Participants’ verbal and gestural responses when explaining the alternative interpretations of the homonym ‘pot’ (in the order in which they were uttered).

<i>Participant</i>	<i>Gesture</i>	<i>Speech</i>
1	-	-
2	-	a. drugs
	-	b. plant pot
3	-	-
4	a. hands create a round space in between them	a. pottery pot

	b. right index and middle finger form a v-shape, moving fore and back in front of the mouth	b. cannabis
5	-	-
6	- -	a. pan b. cannabis
7	- b. right hand imitates to be gripping a handle c. right index and middle finger form a v-shape in front of the face	a. flower pot b. pan c. smoking pot, marijuana
8	a. hands represent a round bowl-shape -	a. physical object b. something that you smoke
9	- -	a. plant pot b. drugs
10	a. a round space is created between the hands -	a. jug thing b. cannabis

A second example stems from a different investigation for which we made some pilot observations. Participants, again, were made to use homonyms to describe individual pictures. For example, one picture showed a desk with a computer, a keyboard and a mouse, some other utensils and a cage with a mouse inside it, playing with its toys. Participants had to refer to both the computer mouse and the animal mouse, and the focus was on how and when they would use speech and gesture. Here, speakers would refer to the computer mouse by holding the right hand in front of the body with the back of the hand pointing upwards, the fingers held together and bent so that they formed a small sphere inside the hand, imitating

the shape a hand adopts when moving a computer mouse. However, this gesture could equally well be used to refer to the animal mouse, showing its shape and size. Interestingly, in this case, speakers tended to distinguish the animal from the PC mouse by referring to things with which the animal was associated in the picture – namely a wheel in which the mouse was running. The accompanying gesture used in this context was that of an extended index finger moving round in quick circles, referring to the wheel's motion.

Although this example also refers to only a few individual instances of gestural behaviour that have been observed, it provides important hints as to what might be happening here. In this last example, it seems that the speakers were aware of the in this case visually shared context between them and their interlocutors. Thus, they were able to draw on the content of the picture as common ground and assume the connection between the wheel and the animal mouse as shared knowledge. Referring to the animal mouse by representing the wheel in which it plays instead of the mouse itself was therefore the most effective way of gesturally disambiguating the two concepts in this particular context.

To sum up, these data of how speakers deal with lexical ambiguity show quite clearly that they use both communicational channels (speech *and* gesture) to resolve ambiguity. Moreover, in instances where requests for clarification are not explicitly posed but potential understanding problems have to be anticipated speakers prevent these from occurring by drawing on the gestural channel also. Furthermore, some individual examples suggest that speakers do not just produce gestures of a 'standardised form' in terms of what they think best represents 'a drinking glass' or a 'cooking pot', irrespective of the context in which a concept is referred to. Rather, it seems that speakers consider what type of information is most disambiguating in the current conversational context, bearing in mind visually shared context as well as the semantic information with which they have provided their addressee in the immediately preceding talk.

However, the above mentioned examples are only first indicators that gestures may be influenced by speakers taking into account what their addressees know and think. The question remains as to whether this is limited to ambiguous speech and to problems in communication, or whether speakers take their addressees' thinking into account on a more general basis. As referred to in the Introduction, people in talk usually share knowledge about the topic of conversation, or they build up shared knowledge over the course of a conversation. This shared knowledge is considered common ground. In talk, speakers do take into account this type of common ground when designing their utterances – at least with regard to the verbal side of utterances; for example, it has been shown that referential descriptions tend to become shorter, generally less complex and reduced to the information required by the addressee to understand the reference (e.g. Clark & Wilkes-Gibbs, 1986). A major question is whether this also affects the gestural side of utterances. If both speech and gesture are part of language, then we should expect that it does. Experimental studies are currently in progress investigating the influence of common ground on speech and gesture use; a first analysis of some of these data is presented subsequently.

3. Study 2

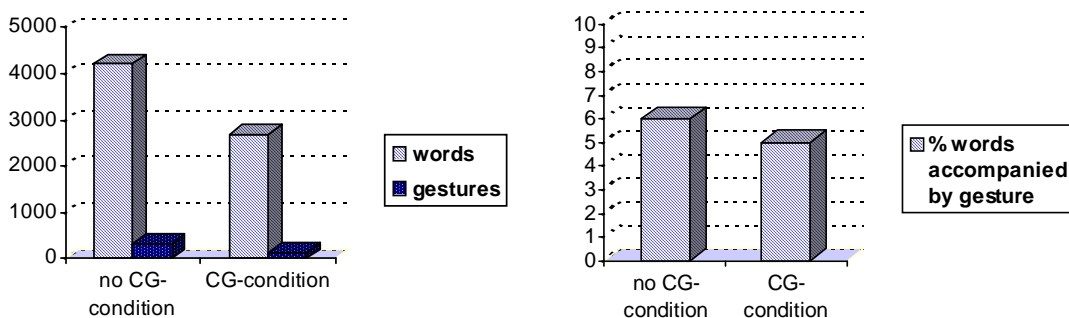
This study experimentally manipulated common ground by using two conditions, one in which pairs of interactants were given the chance to jointly familiarise themselves with the content of a range of stimulus pictures (common ground, or CG-condition), and another in

which participants were not given the opportunity to do so (no CG-condition). There were 8 pairs in each condition which produced data that was considered in the analysis. However, the actual experimental task was the same in both conditions. One participant from each pair was asked to describe the position of a certain entity in each of the picture stimuli. The pictures showed busy scenes of various kinds of objects, such as buildings, as well as cartoon characters carrying out different kinds of actions; the speakers referred to various entities in order to guide their respective addressee, who was not able to see the picture, to the appropriate point in the picture where the target entity was positioned. Based on the speaker's description, the addressee had to mark this position on a copy of the stimulus pictures which were handed to them after each description (but which did not show the target entity).

One aim of this analysis was to find out whether common ground has an effect on how speakers use gesture, or more precisely, whether speakers draw on the gestural channel less often when common ground exists. To test this, the number of words used by speakers in the two conditions was counted as well as the number of iconic gestures. Then the proportional use of gestures was calculated (i.e. number of gestures made, divided by the number of words used) to account for the different lengths of the picture descriptions and thus to arrive at a standardised measure.

The total number of gestures produced in the CG-condition was 130, compared to 318 in the no CG-condition (or an average number of 16.25 compared to 39.75 gestures per speaker). The overall number of words produced in the CG-condition was 2689, compared to 4211 words in the no CG-condition, or an average of 336.13 words per speaker compared to an average of 526.38 words. The proportion of gestures used per a hundred words was 5% (130/2689) in the CG-condition, and 8% (318/4211) in the no CG-condition when considering the *total* number of words and gestures. When calculating the average proportion per speaker, the proportion was 5% in the CG-condition and 6% in the no CG-condition, i.e. in the CG-condition speakers accompanied a mere one per cent less words with gesture. This difference was not statistically significant; ($U=21.5$, $n_1=8$, $n_2=8$, n.s.).

Figs. 1 and 2: Total number of words and gestures produced in the two experimental conditions, as well as the percentage of words accompanied by gesture.

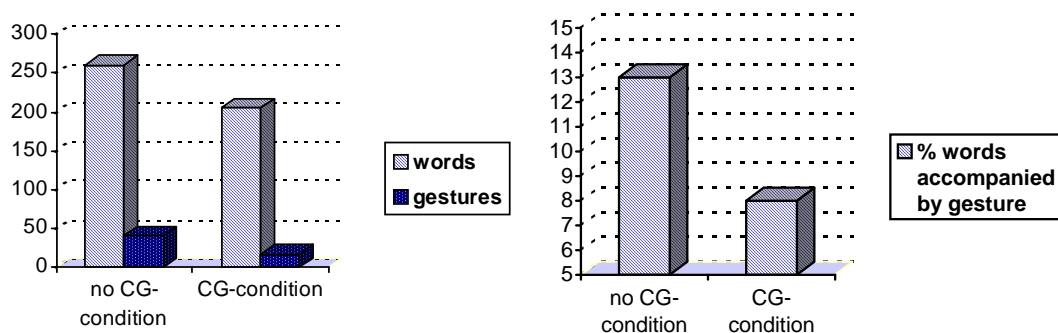


A possible reason for this lack of difference could have been the rather complex stimulus material in that the time participants had to familiarise themselves with the pictures in the CG condition may not have been sufficient for them to take in all of its content, and hence not all of it was assumed as known, thus not considered common ground. For this reason, the same analysis was carried out taking into consideration references to selected entities only (a

house, a bridge, a knot in a pipe), which speakers in both conditions referred to frequently as they were fairly close to the position of the target entity and rather big in the context of the picture, making them very suitable landmarks.

Speakers in the CG-condition used a total of 17 gestures to refer to these entities, or an average of 2.1 gestures per speakers, and speakers in the no CG-condition used a total of 41 gestures when referring to the respective entities, or 5.1 gestures per speaker, on average. The total number of words used to refer to the selected entities in the CG-condition was 205, and the average per speaker was 25.6 words. In the no CG-condition, the total number of words was 261, and the average per speaker was 32.6 words. When considering the *total* number of words and gestures, the proportion of gestures used per a hundred words was 8% (17/205) in the CG-condition, and 16% (41/261) in the no CG-condition (i.e. twice as many gestures were used by speakers in the no CG-condition). When calculating the average proportion per speaker, the proportion was 8% in the CG-condition and 13% in the no CG-condition; however, this difference was not statistically significant ($U=22.5$, $n_1=8$, $n_2=8$, n.s.).

Figs. 3 and 4: Number of words and gestures produced in the two experimental conditions to refer to the selected entities, as well as the percentage of words accompanied by gesture.



The question is whether this lack of significant difference in terms of the proportional use of gestures means that common ground has no effect at all on gesture use. In order to answer this question we have to take a more detailed look at the individual gestural representations. One difference that appeared as rather striking concerned the degree of elaborateness that the gestures showed (by this we mean the degree of definition visible in the gestures), with those from the CG-condition appearing considerably less elaborate. To analyse whether this was a reliable difference, the gestures used to refer to the selected entities were examined more closely. Two independent judges (both blind to the experimental conditions) scored the elaborateness of the 58 individual gestures on a 7-point Likert scale, ranging from ‘very elaborate’ to ‘not very elaborate’. Their scores showed a strong correlation ($r_s(58)=.721$, $p<.0001$). The two scores from the judges were averaged for each gesture to achieve a more objective measure. Based on these scores, an average elaborateness score was determined for each speaker (based on all the gestures a speaker produced with the respective referential descriptions) so that the two experimental groups could be compared statistically. This comparison yielded a significant result ($U=4.5$, $n_1=5$, $n_2=7$, $p<.03$), with the elaborateness of the gestures in the CG-condition being lower than that of the gestures produced in the no CG-condition.

The fact that the proportional number of gestures used by speakers in the two experimental conditions did not differ significantly seems to suggest that the gestures still fulfil an important communicational function even when common ground exists, at least in the context of the experimental task carried out by participants in the study described here. However, the question is what type of function, and whether some of these functions are specific to talk in which common ground exists.

The finding that the gestures produced in the common ground condition were significantly less elaborate than those made in the no common ground condition supports very similar evidence from a study by Gerwing & Bavelas (2004) who found that gestures become more 'sloppy' when common ground exists (which seems to capture something very similar to the 'elaborateness' that we measured). They also found that gestures become significantly less informative when speaker and recipient share common ground. This is a very interesting finding indeed and future research will need to investigate whether the decrease in elaborateness, or precision, affects the representation of semantic information. Further, we need to examine in more detail how gestures become less informative, focusing in particular on how this process influences the semantic interaction of the two modalities, gesture and speech.

4. Conclusion

The findings reported in this paper corroborate previous findings which have shown that social processes in interaction do affect gesture use. Moreover, the findings demonstrate that speakers do anticipate their addressees' thinking when gesturing. This goes against the notion that gestures are not communicatively intended. Further, it shows that gesture production theories need to explicitly incorporate the influence of social processes that are inherent to face-to-face communication. Theories that limit their focus too much on either only the speaker or only the recipient in order to explain the occurrence and use of gestures or their effect on comprehension may not always be looking at the full picture. This argument parallels Clark's (1996) criticism of traditional psycholinguistic theories which focus on either the speaker or the recipient, rather than viewing language use as a collaborative activity between two or more individuals.

References

- Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language & Social Psychology, 19*, 163-194.
- Bavelas, J. B., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture, 2*, 1-17.
- Beattie, G. (2003). *Visible Thought: The New Psychology of Body Language*. London: Routledge.

- Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: An experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, 99, 239-272.
- Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by 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. (2001). An experimental investigation of the role of different types of iconic gesture in communication: a semantic feature approach. *Gesture*, 1, 129-149.
- Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*, 93, 179-192.
- Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: a reply to McNeill. *Psychological Review*, 96, 168-174.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39.
- Furuyama, N. (2000). Gestural interaction between the instructor and the learner in origami instruction. In D. McNeill (Ed.), *Language and Gesture* (pp. 99-117). Cambridge: Cambridge University Press.
- Gerwing, J., & Bavelas, J. B. (2004). Linguistic influences on gesture's form. *Gesture*, 4, 157-195.
- Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?. *Gesture*, 3, 127-154.
- Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal Interaction* (pp. 13-45). Beverly Hills: Sage.
- Kendon, A. (1985). Some uses of gesture. In D. Tannen & M. Saville-Troike (Eds.), *Perspectives on Silence* (pp. 215-234). Norwood: Ablex.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Krauss, R.M., Chen, Y., & Gottesman, R.F. (2000). Lexical gestures and lexical retrieval: A process model. In D. McNeill (Ed.), *Language and Gesture* (pp. 261-283). Cambridge: Cambridge University Press.

- Krauss, R.M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate?. *Journal of Personality and Social Psychology*, *61*, 743-754.
- McNeill, D. (1985). So you think gestures are nonverbal?. *Psychological Review*, *92*, 350-371.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- McNeill, D. (2005). *Gesture & Thought*. Chicago: University of Chicago Press.
- Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and Gesture* (pp. 64-83). Cambridge: Cambridge University Press.
- Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees?. The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*, 688-704.

¹ These figures have previously been published in Holler & Beattie (2003).