Research on Language and Social Interaction, 27(3), 223–237

Communicative Effects of Speech-Mismatched Gestures

David McNeill
Department of Psychology
University of Chicago

Justine Cassell
Department of Computer and Information Science
University of Pennsylvania

Karl-Erik McCullough
Department of Linguistics
University of Chicago

There is increasing agreement that gestures are interesting, even crucial components of language. By studying them in conjunction with speech (not in isolation), we gain new insights into the nature of language itself, how we produce it, how thought and language are related. We discover that languages contain not only words, phrases, and sentences, but languages also have imagery; they have a global, instantaneous noncompositional component that is as defining of the existence of language as are the familiar linguistic components. In this

Preparation of this article was supported by Grant DC–01561 from the National Institute of Deafness and Other Communicative Disorders to the University of Chicago, Grant BNS–8518324 from the National Science Foundation to the University of Chicago, and Grant VPW–GER–9350179 from the National Science Foundation to the University of Pennsylvania.

Correspondence concerning this article should be sent to David McNeill, Department of Psychology, University of Chicago, 5848 University Avenue, Chicago, IL 60637.

article, we argue that speech and gesture work together to convey a single meaning in discourse.

What kinds of gestures do we mean? We focus on the gestures that are the spontaneous and largely unwitting, manual accompaniments to speech. We exclude pantomime, sign language, and emblems. What is meant by *pantomime* is generally interpretable action without accompanying speech. *Sign languages* (of which American Sign Language is a familiar example) are full-fledged linguistic systems, complete with socially established grammars, lexicons, and a historical tradition. Emblems also are descendants of historical traditions and some are extremely long (Kendon, 1981). Many *emblems* occur in the absence of speech, for they can function as language substitutes (rather than as language accompaniments), and they also differ from gesticulations in that they have standards of well-formedness, whereas gesticulations are idiosyncratic in their manner of production.

What Kendon (1980) termed *gesticulations* are the gestures that we study, and they differ from signs, emblems, and pantomime in crucial respects. First of all, with gesticulations speech is nearly always present (better than 90% of such gestures occur during actual speaking; McNeill, 1992). Second, gestures of the gesticulation type are *co-expressive* with speech; that is, they cover the same idea unit as the speech they occur simultaneously with. At the same time, however, they also are *nonredundant,* that is, they supplement speech and have their own communicative effects. Such nonredundant, co-expressive, and synchronous gesticulations form, in the memory of a listener, a single unified system of meaning.

We demonstrate this claim with an experiment carried out on induced speech–gesture (i.e., gesticulation) mismatches.

First, however, a note on the different types of gesture that we have studied. We transcribe gestures into four types that are differentiated semiotically (rather than in terms of movement direction, location, or other behavioral qualities; see McNeill, 1992):

> *Iconics* are representational gestures that display concrete aspects of the scene or event being concurrently described in speech.
>
> *Metaphorics* also are representational gestures but display images of abstract concepts and relationships that typically relate to the concurrent speech on a meta level.
>
> *Beats* mark with baton-like movements words that are significant, not purely for their semantic content, but for their discourse-pragmatic content.

Deictics (abstract) create locations in gesture space for abstract concepts or relationships.

In normal adult discourse, speech and gesture are highly congruent. Gestures have the same semantic and pragmatic content as speech, as in the following example of an iconic gesture:

he grabs a big oak tree and HE BENDS IT way back

[*Right hand appears to grasp and pull back from the front space a large tube-like object.*]

That is, the event of pulling back is presented in speech and in gesture. However, speech and gesture are not redundant. Not every aspect of the event appears in speech and one aspect is more clearly presented in the gesture. The gesture shows that the thing being bent back is attached at one end (it is not like a ruler or wire held in both hands, therefore). The verbal choice of *way back* may index the same implicit anchoring at one end but if so, does it less clearly than does the gesture. The gesture supplements the *bends it way back* clause with explicit information that is, in the most generous interpretation of the speech, implicit. The result of the joint presentation of speech and gesture better fits the previous clause's mentioning of the object being bent, a big oak tree (in this case from a story in a comic book).

As another example, consider the way that left–right space is indexed in narrations. The narrators of an animated cartoon (film) that we have studied virtually never say whether a character or event is on the left or right side of the screen (they mention the vertical but not the left–right dimension); but the narrators know which side events are on, because their gestures invariably show this placement, and do so even when they make use of the opposite hand (the hand crosses over to the other side; McCullough, 1992). Gestures thereby supplement speech.

In this article, we are interested in the communicative effects of this gestural supplementation of speech. We show that the combination of distinct information in gesture and speech has communicative significance and that the listener/viewer does not appear to heed in which channel information is conveyed. Rather, the listener/viewer combines the information of both channels into a single mental representation.

We proceed via an experiment in which mismatches are artificially created and carried to an extreme of separation, mismatches that would not normally occur as part of the process of gesture supplementation.

Consider the following scenarios, one containing a normal gesture and one containing a gesture mismatch:

Matched Gesture

And he's running along ahead of it
[*Hand moves away from the self & towards listener in the upper gesture space as 2 downward pointing fingers wiggle.*]

Mismatched Gesture

And he's climbing up the inside of it
[*Hand moves away from the self & towards listener in the upper gesture space as 2 downward pointing fingers wiggle.*]

In the first case, gesture and speech present more or less the same meanings: There's a character running along something overhead (trolley wires, in the cartoon story our subject was recounting). Even here the gesture is supplementary, conveying the viewpoint of the narrator, who is watching the character run away from him, into the distance; but at least the gesture and speech share the same referent, the event is one and the same. But in the second scenario, gesture and speech clash, they depict different situations: In gesture the character is again running along something overhead, but in speech he is climbing up the inside of something. We are interested in mismatches not for themselves, however, but because of what they reveal about the communicative effect of matching gestures, the kind that do occur naturally. In the experiment to be described, subjects were exposed to a videotape showing a speaker telling a story in which gesture–speech mismatches occurred and then were asked to retell the story from memory. Their retelling of this discourse that they had seen was compared to a similar retelling by other subjects, who only heard an audio recording of the story. If, as we show, the subjects who saw the gesture–speech mismatches retell those parts of the story in which the mismatches occurred differently from the subjects who did not see the mismatches, we can conclude that the mismatched gestures had an effect. From this we can infer that a matching gesture also must have an effect. There is no way that a listener could know he or she is being treated to a mismatch in advance: If we show that mismatches register, then all gestures must therefore register.

THE MISMATCH EXPERIMENT

The mismatch experiment is conducted as follows. The subject sees a videotape of another person telling the story of an animated cartoon; the subject does not watch the cartoon (i.e., in contrast to our standard procedure in many studies, in which the narrator views the actual cartoon and then retells the story from memory to the listener; cf. McNeill, 1992). Unbeknown to the subject, the person on the videotape is one of us performing a carefully choreographed program of mismatching gestures along with a number of normally matching gestures. The subject retells the story to a second subject in the normal manner, without seeing the original cartoon. With this homemade video stimulus, we have found that if subjects see the videotape without a break, they forget a great deal of the detail; consequently, we divide the tape into three parts and have the subjects retell each part before showing them the next. For the details of the experiment, the procedure, the subjects, and several versions of the mismatch tape, see Cassell, McNeill, and McCullough (1994). What we hope to observe with this procedure are the traces in the subject's own narration of the mismatched speech–gesture combinations that we have planted into the video narrative.

The Mismatches

We introduced three kinds of mismatches (several occurrences of each) that varied in how radically they diverge from speech; the first category is fairly typical of the ways gesture supplements language, and the other two truly oppose the content of the accompanying speech.

Manner Mismatches

These provide additional or different information about the manner of motion that in speech is conveyed with an unmarked motion verb. This is the one case where we could say the mismatch has the same referent as the verb, and more correctly call it a nonredundancy. An example is the following:

He goes up the pipe
Hands rise upward
moving hand over hand.

The verb encodes ascent and motion in relation to a deictic center (Buhler, 1965/1934) but not the manner of ascent. This is shown in the gesture to be climbing as if on a ladder. Manner mismatches can occur in spontaneous gesticulation.

Perspective Mismatches

Here, a story character is presented from a different spatial perspective without a corresponding new theme in speech. In normal discourse, new gestural perspectives signal thematic discontinuities (McNeill & Levy, 1993). To shift perspective without reference to a new theme is a mismatch more radical than the manner mismatch just described, and such mismatches are rarely if ever seen in spontaneous discourse. An example is the following:

Granny sees him and says oh, what a nice little monkey"
[*RH points into space in front of self*
and this space now represents Sylvester.]

And then she offers him a penny
[*RH moves as if giving*
a penny to own body,
which here represents Sylvester.]

During the first gesture, the perspective places Sylvester in front of the speaker, but during the second gesture, the perspective shifts and Sylvester appears to be at the speaker's own locus. Because there is nothing in speech to signal thematic discontinuity, this is a mismatch.

Anaphor Mismatches

These are shifts in space over time without an accompanying linguistic shift or discontinuity. These, too, are more radical mismatches

than normal gesture supplementation. The narrator sets up contrasting referents in space, one to the left and one to the right and then violates this positioning by carrying out an action by one of the referents in the other referent's space. In this example, the *he* in *he lunges* means Sylvester but the hand to depict the lunging is the one that had represented Tweety:

Sylvester is right near (Tweety), watching him,
[*LH represents Sylvester, RH represents Tweety*]

And then suddenly he lunges for him and runs into the apartment after him
[*RH, now Sylvester, lunges forward.*]

Because the continuation of reference to Sylvester with the pronoun *he* implies continuity, the spatial shift is a mismatch.

Thus, we have a range of mismatches: from an approximation to normal gesture supplementation in the manner mismatches, to more radical partings of the ways in the perspective and anaphor mismatches.

If gestures have an effect on the listener, the mismatches should also have an effect, and, because they are unusual, will be apparent to us in coding from the other gestures that would normally appear without a mismatch. For example, an anaphor mismatch may trigger a reference error that would never otherwise occur. A perspective mismatch may set off a restructuring of the gesture space in a way not normally seen; a manner mismatch may cause the speaker to invent an event different from any presented in the original narrative, again an easily spotted departure from the normal performances of subjects.

In addition to the main subjects who saw the videotaped narration and were exposed to mismatches, control subjects were presented with only the soundtrack. For these subjects, no mismatches could occur, but they were exposed to the same words, prosody, and timing. This is a control to show that gestures have effects. If gestures have no effects on the listener's uptake of information, mismatches also will have no effects and the subjects who were shown the video should be the same as those who heard the soundtrack.

Two coders examined each subject's narration. The coders, who were blind to the version (audio and video, or audio only) of the narration that the subjects had been exposed to, did not know when they coded the uptake of a mismatch whether that subject had actually seen

a mismatch. To code the narratives, the coders relied on very clear expectations of what kinds of gestures and narrative speech ordinarily take place in retelling this particular cartoon story (i.e., the cartoon has been shown to many subjects under normal viewing conditions; McNeill, 1992). The coders could examine each experimental subject's narrative for unusual gestures, unusual linguistic phenomena, invented scenes, and omissions of scenes.

The Effects

The following shows the percentage of mismatches producing some detectable effect in these ways on the narrative performance of the 12 subjects in the test by Cassell et al. (1994).

Anaphor	30%
Manner	52%
Spatial	46%

Even more telling are the overall results of the task. There were a total of 117 potential speech-gesture mismatches to which the 12 subjects were exposed. There were a total of 66 speech-gesture pairs coded as uptake of a mismatch. A null hypothesis would be that these 66 are equally divided between the actual mismatch condition and the audio only condition. Another null hypothesis is that they are found as often in cases where no mismatch was viewed as when one was viewed. If either of these null hypotheses was the case, then we merely discover something about our coding procedure. However, if the 66 were primarily found in actual cases of mismatch, then we may appropriately claim that the mismatches have an effect on listeners' understanding of the story. The latter turned out to be the case: 52 codes of "uptake" occurred in actual mismatch situations, whereas just 7 codes occurred in the audio-only condition and 7 codes occurred in places where a normal gesture, but not a mismatch, was seen. In the view of these results, we claim that listeners are influenced by speakers' gestures.

In terms of where the effect of a mismatch was located, Cassell et al. (1994) found nearly the same degree of effect on the subjects' gestures and their speech. The following table shows how often changes

that were attributed to mismatches appeared in speech, in gesture, and in both speech and gesture.

Speech	23%
Gesture	19%
Speech and Gesture	10%

Thus, the effects of mismatching gestures were felt not in gesture alone but also in speech, and in speech and gesture simultaneously (that the effects appear in both speech and gesture about half as often as in either channel alone, is not surprising). We interpreted this effect of mismatches spreading evenly across the subjects' speech and gestures as evidence that subjects, when they are exposed to a gesture–speech mismatch, do not code the information according to the channel of input. Rather, the subject retains information in memory in neutral form, not indexed either as speech information or gesture information.

If listeners cannot resolve the mismatch between speech and gesture, they may choose to simply omit that part of the narration from their retelling. We, therefore, looked at omission of events in the presence of mismatches and normal gestures. The results were inconclusive but suggestive: There were 18 omissions of events at potential mismatch points and 11 of these followed actual mismatches (7 after straight presentations).

Thus, we found tracers of mismatches in the speech, the gestures, and the omissions of events, from which we concluded that gestures form a part of the communicative message that listeners receive and mismatching gesture-and-speech inputs create communicative and mnemonic problems for the listener.

ILLUSTRATIVE MISMATCHES

A Referential Error Following an Anaphor Mismatch

The mismatch in this case was due to a combination of a cohesive pronoun use with a shifting of the gesture space. The pronoun spoke of

cohesion, the space of referential change. The subject temporarily was seduced by the gesture, made a referential error, then repaired her mistake. In speech the on-screen narrator said, *So he goes up the pipe.* [*Tweety's*]$_1$ *singing and swinging in his window,* [*and Sylvester's*]$_2$ *right near him watching him and then suddenly he* [lunges]$_3$ *for him and runs into the apartment after him,* where the first gesture seemed to hold Tweety in the upper right location, the second gesture seemed to hold Sylvester in the upper left location, and the third gesture (the mismatch) was the right hand (which had been Tweety) lunging into the right space, whereas the linguistic signal, *he lunges,* continued to refer to Sylvester. The subject recalling this episode made and immediately repaired a reference error: *and then Tweety goes down and tries to climb—I mean Sylvester goes down and tries to climb* (the subject also ran together this singing-and-swinging scene with the scene immediately following in the cartoon). Thus, in this example, the effect of the mismatch was realized in successive referential choices: first, the one driven by space (Tweety the actor) and, second, the one driven by speech (Sylvester the actor), which was of course the correct choice.

A Lexical and Spatial Repair Following a Perspective Mismatch

The mismatch was, first, a gesture by the on-screen narrator that seemed to establish the speaker's own locus as the origin of a depicted action, followed immediately by a second gesture that showed a continuing action by the same character but from a locus in front of the speaker. Speech contained no signal of a change of perspective; on the contrary, speech signaled continuity. The subject in this case experienced a surprising transformation of both gesture and language in attempting to retell this scene. (This example was first noted and analyzed by Dray & McNeill, 1990.) In speech, the on-screen narrator said, [*then he steals his costume*]$_1$, *then Sylvester* [*dresses up*]$_2$ *in the monkey's costume,* where the narrator's first gesture showed pulling the costume away from his own locus and the second gesture showed putting the costume on a character in front of the speaker. The subject spoke as follows: *and lures the* [*mon*]$_3$*key over to him, and kidnaps* [*or kidnap*]$_4$ [*hijacks him and basically*]$_5$ *takes his outfit . . . ,* where Gestures 3 and 4 established the monkey in the space in front of the

speaker and Sylvester at her own locus (in Gesture 3 she pointed down to her own lap; in Gesture 4 she reached out into the front space); then the subject abruptly recast this arrangement. Gesture 5 was performed by grasping her own shoulders, thus placing the monkey at her locus and Sylvester at the front, whereas the Gesture 4 had placed the monkey in the space before her. At the same time, she altered her verb from *kidnap* to *hijack* and the verb shift is completely parallel to the gesture shift. Where *kidnap* implies an intruder absconding from the place, *hijack* focuses on the intruder entering into it. In this way, the subject totally negated the mismatch in both speech and gesture. Dray and McNeill (1990) attributed all of this rearrangement to the indigestible quality of the on-screen mismatch. This example differs from the previous example in which the mismatch triggered successive referential choices; here the mismatch set off a total overhaul of language and gesture.

A Wholly New Scenario After a Manner Mismatch

The gesture of this example conveys, in the on-screen narration, a manner of motion (bouncing) and is paired with a verb that does not indicate manner (*comes out,* which indexes deictic perspective but is compatible with a variety of manners). The gesture returns in the subject's narration, not as a gesture, however, but as a new scene that had not been in the original narrative at all (*goes down stairs*). Here the mismatch was resolved by inventing a new scenario, the character negotiating stairs that had not been shown or told before. The on-screen narrator had said, [*he comes out the bottom of the pipe*], bouncing his hand up and down. The subject turned this into *and* [*then goes down stairs*] *across—back across into . . .*, with a neutral gesture (the hand dropping straight down).

We see in this example an instance of our conclusion, that stored information is not indexed according to its channel of input. This we see in the fact that the bouncy input gesture re-emerged in the subject's memory as a false image of stairs while her own gesture was a simple drop motion. Thus, the bouncy image was stored in some general form that no longer included its gestural source and was actually more available to the subject for a linguistic invention than to be re-externalized as a gesture.

We see in the preceding examples, how the communicative effects

of gestures vary across a range in which the consequences for memory are increasingly significant. At a minimum, the subject performs successive encodings of the two components of the mismatch but remains capable of recognizing which is the appropriate one. A more serious distortion is the subject carrying out a perspective rearrangement of language and gesture. The most extreme distortion, and stepping over the line into a false memory, is the subject setting up a new scene.

THEORETICAL SIGNIFICANCE: HOW DOES MISMATCH WORK?

What framework can we adopt to interpret the effects of speech–gesture mismatches? We take as a clue the evidence that information, once absorbed, is not indexed by the input channel and that subjects, after exposure to a mismatch, attempt to resolve the mismatch in their own retelling. Subjects, in other words, do not try to re-create the input to which they were exposed, but they try to form a coherent mental model and introduce changes in memory, where it is necessary to make this possible. We propose that subjects are skilled at these transformations and carry them out without awareness because gesture–speech nonredundancies (of which artificial mismatches are an extreme form) are normally part of the process of communication. Different information in the two channels is not simply slippage in the speech–gesture system but an operative part of communication. The principal raison d'être of a nonredundancy is to fuel cognitive change in the listener: an undeniable good in communication. That is, getting the interlocutor to understand and to see something in a new way, may be particularly fostered by gesture–speech nonredundancy. Therefore, adaptation to the presence of different information in speech and gesture and having strategies for combining information from the two channels are well developed in adult subjects.

That nonredundancy may play a role in inducing cognitive change has been proposed for children's conceptual development by Goldin-Meadow, Alibali, and Church (1993). Goldin-Meadow et al. found that, at a certain stage of development, children regularly display different strategies in speech and gesture during problem solving tasks. In one of

their experiments, children were given problems in arithmetic such as the following: 6 + 7 + 4 = ____ + 4. Eight- and 9-year-old American children find this kind of problem difficult. They may not solve the problem because they add all the numbers to the left of "=" and put this sum into the blank. However, the child often makes a simultaneous gesture that displays a more insightful strategy—for example, the child points to the first two numbers and to the blank in order to show that elements of the solution were being taken into account on the gesture level if not on the verbal level. The surprising discovery by Goldin-Meadow et al. was that children with such mismatching gestures and speech give independent evidence of being in a transitional knowledge state. They were more likely to learn from minimal training, for example. Also, in an experimental paradigm in which cognitive load was measured during arithmetic problems, on trials in which children happened to exhibit speech–gesture mismatches, they had greater cognitive loads compared to trials when the gestural and spoken strategies were the same, implying that more information was being handled on the mismatch trials.

We think that a similar mechanism may also be at work in normal discourse, fueling an understanding of language that goes beyond the speech uttered. In this case, the process would operate on-line, keeping up with the flow of discourse, and would involve nonredundant speech–gesture combinations and their resolution. Its function would be the adult's on-line counterpart of conceptual change: new understanding by the listener.

The following shows how gesture and speech can combine into a single system of meaning (see McNeill, 1992, for extensive discussion). The listener normally takes in both gesture and speech. This is part of an interaction of image and word involved in linguistic processes in general, and it is done without the necessity of conscious attention; the two channels smoothly combine into a single idea unit. The *bends it way back* example displayed earlier presented a harmonious blend of words and gesture, each channel covering some of the content presented by the other, while presenting other content different from the other. The different contents meshed into a coherent picture of the scene of the character bending back something fastened down at just one end. This is the sense in which two nonredundant channels can be said to cover the same idea unit, the two combining into a single idea unit about the scene that is richer than the picture conveyed by either channel alone. The

stairs example demonstrates that gesture and speech information are stored without specific indexing of the channel of input; the information from the two channels combines into neutral form equally available to speech or gesture.

The important insight we have to offer is that when something unusual happens, as in our experiment with artificial mismatches, the listener still attempts to form a single idea unit, an integrated combination of speech and gesture. However, now the combination is only possible with further processing. We saw in the preceding examples some of the extra processing steps that may take place—successive ideas, reorganization of the idea unit, or fabrication of a new one; and no doubt others could be found. The point we wish to emphasize in the involuntary, automatic character of forming an idea unit out of information from the two channels, even when gesture and speech do not directly fall together, as in our experiment. When the nonredundancy is within the normal range, this process would operate no less smoothly.

Such a process in on-line understanding emphasizes a dialectic in cognition: incoming information from two media, two semantic systems in contact, and possibly in conflict. The importance of nonredundancy of speech and gesture is that a new idea unit may be formed to resolve the conflict, and in this way a mismatch sets the stage for a new form of understanding and, thus, is worth extra processing steps.

The phenomenon of gesture supplementation can thus be seen stretching along a continuum with gesture mismatch at one extreme. This continuum implies that the resolving of two kinds of information may be a constant factor of communication, that mismatch is a regular mechanism of the process of communication. Thus, we propose that the combining of gestures with language is part of the process of communication both in production and in comprehension.

The experiment on induced speech-gesture mismatches exposed this essential process in a baroque manner, but it did not create anything that listeners could not accommodate.

REFERENCES

Buhler, K. (1965). *Sprachtheorie: Die Darstellungsfunktion der Sprache* [*Language theory: The representational function of language*]. Stuttgart: Gustav Fischer Verlag. (Original work published 1934)

Cassell, J., McNeill, D., & McCullough, K.-E. (1994). *Speech–gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information.* Manuscript submitted for publication.

Dray, N. L., & McNeill, D. (1990). Gestures during discourse: The contextual structuring of thought. In S. L. Tsohatzidis (Ed.), *Meanings and prototypes: Studies in linguistic categorization* (pp. 465–487). London: Routledge.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review, 100,* 279–297.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relation between verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.

Kendon, A. (1981). Geography of gesture. *Semiotica, 37,* 129–163.

McCullough, K.-E. (1992, November). *Visual imagery in language and gesture.* Paper presented at the annual meeting of the Belgian Linguistic Society, Antwerp, Belgium.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

McNeill, D., & Levy, E. T. (1993). Cohesion in speech and gesture. *Discourse Processes, 16,* 363–386.