

# **Gesture and the process of speech production: We think, therefore we gesture**

Martha W. Alibali

*University of Wisconsin–Madison, Madison, USA*

Sotaro Kita

*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Amanda J. Young

*New York University, New York, USA*

At what point in the process of speech production is gesture involved? According to the Lexical Retrieval Hypothesis, gesture is involved in *generating the surface forms* of utterances. Specifically, gesture facilitates access to items in the mental lexicon. According to the Information Packaging Hypothesis, gesture is involved in the *conceptual planning* of messages. Specifically, gesture helps speakers to “package” spatial information into verbalisable units. We tested these hypotheses in 5-year-old children, using two tasks that required comparable lexical access, but different information packaging. In the *explanation* task, children explained

---

Requests for reprints should be sent to Martha W. Alibali, Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53706, or Sotaro Kita, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands. Email: mwalibali@facstaff.wisc.edu or kita@mpi.nl.

Martha W. Alibali, Department of Psychology; Sotaro Kita, Language and Cognition Group; Amanda J. Young, Shirley M. Ehrenkranz Graduate School of Social Work.

This paper is based on a Senior Honors Project conducted at Carnegie Mellon University by Amanda J. Young, under the supervision of Martha W. Alibali and Sotaro Kita. The research was supported by a grant from the Undergraduate Research Initiative at Carnegie Mellon University to Amanda J. Young, and by the Max Planck Institute for Psycholinguistics. We thank Tracy Barrett for assistance with data collection; Andrea Krott for assistance with the analysis based on the CELEX lexical database; and Nicole McNeil, Asli Özyürek, and Jan-Peter de Ruiter for helpful discussions and constructive feedback. We are especially grateful to Sharon Carver, Jean Simpson, and the teachers and students at the Carnegie Mellon University Children’s School, who made this work possible.

why two items did or did not have the same quantity (Piagetian conservation). In the *description* task, children described how two items looked different. Children provided comparable verbal responses across tasks; thus, lexical access was comparable. However, the demands for information packaging differed. Participants' gestures also differed across the tasks. In the explanation task, children produced more gestures that conveyed perceptual dimensions of the objects, and more gestures that conveyed information that differed from the accompanying speech. The results suggest that gesture is involved in the conceptual planning of speech.

When people speak, they often spontaneously produce gestures, and those gestures are typically used to indicate or represent objects and ideas (McNeill, 1992). Many studies have shown that gestures play a role in communicating information to listeners (see Kendon, 1994, for a review). Other studies have suggested that gestures also serve a purpose for speakers (e.g., de Ruiter, 1998; Krauss, Chen, & Chawla, 1996). In this paper, we focus on the functional role of gesture in the process of speech production. Our goal is to test two alternative accounts about how and when gesture is involved in speech production. The two accounts make different claims about the level of conceptual or linguistic processing in which gesture is involved.

One account, which we call the Lexical Retrieval Hypothesis, holds that gesture is involved in *generating the surface forms* of utterances. More specifically, gesture plays an active role in lexical access, particularly for words with spatial content (Butterworth & Hadar, 1989; Krauss, 1998; Rauscher, Krauss, & Chen, 1996). One version of this view holds that iconic gestures derive from spatially encoded knowledge, and facilitate access to lexical entries that incorporate syntactic and semantic information (e.g., Krauss et al., 1996). An alternative version holds that iconic gestures derive from lexical entries (Butterworth & Hadar, 1989) or from non-propositional representations in working memory (Krauss, Chen, & Gottesman, in press), and assist in the retrieval of the relevant phonological forms. By both accounts, gestures serve to boost activation levels for the to-be-retrieved items. Thus, according to the Lexical Retrieval Hypothesis (both versions), gesture plays a role in generating the surface forms of utterances. In brief, according to the Lexical Retrieval Hypothesis, gesture plays a direct role in the process of *speaking*.

An alternative account, the Information Packaging Hypothesis, expands on two ideas drawn from McNeill (1992): first, that “[g]estures, together with language, *help constitute thought*” (p. 245), and second, that gestures reflect the imagistic mental representation that is activated at the moment of speaking. The Information Packaging Hypothesis holds that gesture is

involved in the *conceptual planning* of the message to be verbalised, in that gesture helps speakers to “package” spatial information into units appropriate for verbalisation (Kita, in press). More specifically, speakers use gesture to explore alternative ways of encoding and organising spatial and perceptual information. Thus, gesture plays a role in speech production because it plays a role in the process of conceptualisation. This opens the possibility that gesture may play a role, not only in speech production, but also in other cognitive activities, such as reasoning and problem solving (see Alibali & DiRusso, 1999). In brief, according to the Information Packaging Hypothesis, gesture plays a role in *thinking*.

In this study, our goal is to test these two accounts of the role of gesture in speaking. Specifically, we seek to establish whether gesture is involved in the conceptual planning of messages, or whether it is involved only in the generation of the surface forms of utterances. To accomplish this goal, we developed two tasks that elicit comparable utterances, but make different demands on the conceptual “packaging” of spatial information. If gesture serves to facilitate generation of the surface forms of utterances, as predicted by the Lexical Retrieval Hypothesis, then gesture production in the two tasks should be comparable. Any differences in gesture production across tasks would provide evidence against the Lexical Retrieval Hypothesis. If gesture plays a role in conceptualisation, as predicted by the Information Packaging Hypothesis, then gesture production should vary across the tasks in specific ways. As we describe below, based on an analysis of the two tasks, we derived specific predictions about *how* gesture production should differ across tasks under the Information Packaging Hypothesis. Certain patterns of differences across tasks would provide specific support for the Information Packaging Hypothesis.

For one of the tasks, we chose a classic task from the developmental literature: Piagetian conservation. In the standard conservation task, children judge whether two equal quantities of some item (e.g., two glasses of sand) are the same or different after one has been transformed in some way (e.g., one glass of sand has been poured into a bowl). As one of the tasks in the present study, we asked children to *explain* their conservation judgements. We also developed a related task for which we expected children to provide verbal responses similar to those they provide in the conservation explanation task, but for different reasons. In the second task, we asked children simply to *describe* how two items (e.g., a glass of sand and a bowl of sand) looked different.

We expected children to focus on perceptual dimensions of the task objects in *both* tasks. In the conservation explanation task, young children (particularly those who fail to conserve) often focus on perceptual dimensions of the task objects, as many studies have shown (e.g., Piaget, 1967). In the description task, children are directly asked to describe how

the objects look, and thus it seemed very likely that children would focus on perceptual dimensions of the task objects. Indeed, as expected, children in the present study often provided comparable verbal responses in both tasks (e.g., “This one’s tall and this one’s short”). If comparable lexical items are accessed in the two tasks, then under the Lexical Retrieval Hypothesis, gesture production should not differ across the tasks.

Under the Information Packaging Hypothesis, if there are differences in conceptualisation for speaking, then gesture production should differ across tasks. As we describe below, we argue that there are two important differences in conceptualisation across the two tasks: (1) children’s thinking is more *constrained* in the explanation task than in the description task, and (2) children’s thinking is more *complex* in the explanation task than in the description task. Based on these differences in conceptualisation, one can derive specific predictions about *how* gesture production should differ across tasks under the Information Packaging Hypothesis.

First, we argue that children’s thinking is more *constrained* in the explanation task than in the description task. In the explanation task, but not the description task, children must decide whether the two quantities are the same or different amounts, and therefore they must identify the *particular* dimension(s) that are relevant to the quantity comparison. That is, children must zero in on the particular dimension(s) (e.g., height, width, shape, etc.) that justify their judgements about whether the quantities are the same or different. In contrast, in the description task, children simply focus on how the two objects *look* different. Consequently, *any* dimension along which the task objects differ in appearance is an adequate response, and children need not focus on a dimension that is relevant to a judgement about quantity. Thus, children’s responses are more *constrained* in the explanation task than in the description task.

How should gesture differ across tasks if conceptualisation for one task is more constrained than conceptualisation for the other? According to the Information Packaging Hypothesis, in the process of zeroing in on the relevant dimensions, children will use gesture to explore various dimensions of the task objects. Thus, the explanation task should elicit more gestures representing physical properties of the task objects than the description task. Further, children may explore multiple dimensions of the task objects, but may ultimately express only a subset of these dimensions (those relevant to the quantity comparison) in speech. Put another way, not all of the dimensions that children explore will “survive” into their final utterances. Thus, the explanation task should elicit more gestures that convey information not expressed in speech. Such non-redundant gestures (sometimes called “gesture-speech mismatches”) have been described in previous studies of children acquiring the concept of conservation (Church & Goldin-Meadow, 1986).

Second, we argue that children's thinking is more *complex* in the explanation task than in the description task. Children often solve conservation tasks correctly by considering multiple dimensions of the task objects (e.g., the height and width of a container, as in "This one's taller, but it's also skinnier"). Thus, in the explanation task, we expected children to sometimes consider multiple perceptual dimensions in justifying their judgements. In contrast, in the description task, we expected children to rarely consider multiple dimensions in responding to each problem, because they need not do so in order to solve the task correctly. Thus, the thinking leading up to the response is likely to be less complex in the description task than in the explanation task.

How should gesture differ across tasks if conceptualisation for one task is more complex than conceptualisation for the other? According to the Information Packaging Hypothesis, children use gesture to explore multiple dimensions of the task objects. However, as noted above, in some cases, they may express only a subset of these dimensions in speech. Indeed, previous studies have shown that children on the verge of acquiring conservation sometimes express some perceptual dimensions in speech and others in gesture (Church & Goldin-Meadow, 1986). Thus, under the Information Packaging Hypothesis, the explanation task should elicit more gestures that convey information not expressed in speech.

Note that, in both tasks, children need to refer to the task objects, and they almost always use deictic referring terms such as "this one" and "that one" to do so. Such expressions *require* a deictic (indicating) gesture in order to be interpreted by the listener, and in this sense, the gestures produced along which such deictic terms serve a communicative function. However, children can use gesture to refer to the task objects in different ways. In some cases, children produce gestures that serve *only* to indicate the task objects (e.g., saying "this glass is wide" while pointing to the side of the glass). In such cases, children's gestures are purely deictic. In other cases, children produce gestures that serve *both* to indicate the objects, and to represent or highlight some physical properties of the objects (e.g., saying "this glass is skinny" while holding both hands at the sides of the glass to show the width of the glass). In such cases, children's gestures have both deictic and representational components.

The explanation and description tasks used in the present study are similar in terms of the need for deictic references to the task objects. Thus, we expected children to produce gestures for deictic references in both tasks. Hence, we do not predict a difference in the overall *rate* of gesture production across tasks, under either hypothesis. However, under the Information Packaging Hypothesis, we predict a difference in the *types* of gestures produced in the two tasks. As described above, we expected children to produce more gestures that represent physical properties of the

task objects in the explanation task. Thus, in the explanation task, we expected children to often produce deictic gestures that incorporate a representational component. In the description task, we expected children to often produce deictic gestures that do *not* incorporate a representational component.

Let us briefly summarise the predictions. First, if gesture is involved only in the generation of the surface forms of utterances, as claimed by the Lexical Retrieval Hypothesis, then gesture production in the two tasks should be comparable. If gesture also plays a role in conceptual planning, as claimed by the Information Packaging Hypothesis, then gesture production in the two tasks should differ in two specific ways. First, in the explanation task, children should use representational gestures to explore dimensions of the task objects and to help zero in on the dimensions relevant to the task solution. Thus, children should produce *more representational gestures* (i.e., gestures that represent dimensions of the task objects, such as height, width, shape, etc.) in the explanation task than in the description task. Second, in the explanation task, children should be especially likely to consider multiple dimensions of the task objects, both in the course of zeroing in on the relevant dimensions, and in solving the problem correctly (i.e., by identifying dimensions that compensate for one another). When children consider multiple dimensions of the task objects, they sometimes express some dimensions in speech and others in gesture. Thus, children should produce *more non-redundant gestures* in the explanation task than in the description task.

## METHOD

### Participants

Twenty-two kindergartners from a university laboratory school participated in the study. One boy was excluded because he was unable to stay focused on the experimental task, and one boy was unavailable for testing for the second experimental session. Two girls were excluded because of experimenter error. Thus, there were a total of 18 children (14 boys and 8 girls; mean age 5;6 range 4;11–6;2) who completed the experiment.

### Procedure

Participants completed two tasks in two separate experimental sessions, held approximately two weeks apart ( $M = 15$  days). The order of the two tasks was counterbalanced, and the order of problems within each task was randomised.

*Problem explanation task.* In the problem explanation task, participants were asked to solve and explain a set of six Piagetian conservation problems. They included three continuous quantity problems (conducted using sand) and three mass problems (conducted using playdough). Each trial consisted of two phases. First, participants were asked to verify that two quantities (i.e., two glasses of sand or two balls of playdough) were the same amount (e.g., “Do these two glasses have the same amount of sand, or different amounts of sand?”). One quantity was then transformed (sand poured into another glass, or playdough shaped into another form). After the transformation, participants were asked to judge whether the two quantities were the same or different (e.g., “Is there more sand in this glass (indicating base of one glass), more sand in this glass (indicating base of other glass) or the same amount in both?”) and to explain their judgements (e.g., “How can you tell?”).

*Description task.* In the description task, participants were presented with six tasks comparable to those used in the explanation task, using the same task objects. However, instead of comparing the quantities, participants were simply asked to describe how the two items (glasses of sand or masses of playdough) looked different (e.g., “How do these two look different?”).

### Identifying pairs of responses for coding and analysis

Children’s verbal responses to both tasks were transcribed and divided into clauses, each of which contained one semantic verb. Although many of the children’s responses were similar across tasks, we found that there were some types of responses that were produced only in the explanation task (as we expected based on previous studies of children’s reasoning about conservation, e.g., Siegler, 1995). Specifically, in the explanation task, children sometimes talked about the transformation (e.g., “because you moved it”) or about the identity of the objects (e.g., “You didn’t take any playdough away”). Children never provided these types of responses in the description task. Therefore, to ensure that lexical access was similar across the two tasks, we limited our analysis to pairs in which the responses to both tasks described *dimensions* of the task objects (e.g., width, height, shape, etc.).

We also excluded pairs of responses in which children did not produce gestures in one or the other of the two tasks. Such responses tended to occur in the explanation task because of its discourse context. In the explanation task, when initially asked whether the two quantities were the

same or different amounts, children often judged one quantity to be more (saying, e.g., “this one’s more”). In such cases, when asked to explain their judgement, children sometimes provided extremely brief, single-clause explanations referring only to that object (e.g., “Because it’s bigger”). In such cases, it is not necessary to gesturally indicate the object during the explanation, and indeed, such brief explanations were often produced without any gestures at all. In contrast, in the description task, children almost always provided a full comparative explanation that focused on each object in turn (e.g., “This one’s bigger and this one’s smaller”). Consequently, they almost never produced responses without gesture in the description task. Because our goal was to explore the level of speech production processes in which gesture is involved, we limited our analysis to responses that included gesture, and we evaluated the nature of the gestures produced in each task.

Thus, the final sample for study consisted of 106 responses (53 pairs). Both members of each pair of responses focused on dimensions of the task objects, and both members of each pair included gesture. The responses in the final sample were drawn from 14 different children, with an average of 4.4 pairs of responses per child.

### Coding children’s responses

The content of children’s speech and gestures was coded, both at the response level and at the individual gesture level. The response-level coding was completed first. Children’s verbal and gestured responses were coded on two separate passes through the videotape, following the procedure described by Church and Goldin-Meadow (1986) for coding children’s spoken and gestured explanations of conservation tasks. Verbal responses were coded without viewing the video portion of the tape, and gestured responses were coded without listening to the audio portion of the videotape. After the response-level coding was completed, the individual-gesture-level coding was carried out on an additional pass through the tape, relying on both the audio and the video, and utilising the content codes derived from the response-level coding. This procedure is described in detail below.

### Response-level coding

*Coding verbal explanations.* The content of children’s verbal responses was coded using the system developed by Church and Goldin-Meadow (1986). Examples are presented in Table 1.

*Coding gestured explanations.* The content of children’s gestured responses was also coded using the system developed by Church and



TABLE 1  
Response-level coding of spoken and gestured responses

Speech examples	
<i>Playdough tasks</i>	
	“Because it’s all flat.” ( <i>Shape of transformed playdough</i> )
	“That one’s bigger and that one’s smaller.” ( <i>Size of both masses of playdough</i> )
<i>Sand tasks</i>	
	“This is just really fat.” ( <i>Width of transformed glass</i> )
	“This one is lower, and this one is higher.” ( <i>Level of sand of both glasses</i> )
Gesture examples	
<i>Playdough tasks</i>	
	RH flat palm held over flattened playdough ( <i>Shape of transformed playdough</i> )
	RH point to transformed playdough, then to untransformed playdough ( <i>Indicate both masses of playdough</i> )
<i>Sand tasks</i>	
	LH “C” handshape around base of transformed glass ( <i>Width of transformed glass</i> )
	RH point to level of sand in transformed glass, then to level of sand in untransformed glass ( <i>Level of sand of both glasses</i> )

Goldin-Meadow (1986).<sup>1</sup> The stream of manual behaviour was segmented into individual gestures (see McNeill, 1992), and for each individual gesture, the handshape, placement, and motion were described. Next, a meaning was assigned to each gesture, using Church and Goldin-Meadow’s system. Finally, the content of each gestured response as a whole was inferred based on the meanings of the individual gestures that made up the gestured response. All gestures were initially transcribed and coded by a single coder, and then all codes were checked by a second coder. Examples are presented in Table 1.

*Coding the relationship between gesture and speech.* After the verbal and gestured explanations were coded at the response level, the relationship between gesture and speech was assessed for each response, following the principles laid out by Church and Goldin-Meadow (1986). Any response in which the gestured explanation contained information not expressed in the verbal explanation was classified as a *gesture-speech mismatch*. Any response in which all of the information expressed in the

<sup>1</sup> Unlike Church and Goldin-Meadow (1986), we considered manipulations of the task objects to be gestures if they were produced along with speech, and if they were integrated with speech as gestures typically are (McNeill, 1992). For example, if a child picked up one of the glasses of sand while saying “This cup is big”, the child’s action of picking up of the glass was considered an indicating gesture. We based this decision on the view that such functional actions are often part of communicative acts (Clark, 1996; Engle, 1998).

gestured explanation was also expressed in the verbal explanation was classified as a gesture-speech *match*.

We further subdivided gesture-speech match responses into two types. Any gesture-speech match response in which all gestures were simple indications of the task objects was classified as an *indication-only match* response. Any gesture-speech match response in which at least one gesture conveyed substantive information about the task objects was classified as a *substantive match* response. Examples of responses of each type are presented in Table 2.

### Coding individual gestures

Each individual gesture in each response was then classified as either a *substantive* gesture or a *simple indication*. We defined a substantive gesture as a gesture that conveys some dimension of the task objects by virtue of its handshape, placement, or motion (e.g., depicting the *width* of one of the glasses of sand with a gesture using a C-shaped handshape, indicating the *level* of the sand in one of the glasses by pointing to the level on the side of the glass, tracing the round *shape* of the flattened ball of playdough). We defined a *simple indication* as a gesture that simply indicates one or the other of the task objects but does not convey a particular aspect or dimension of the object (e.g., pointing to one of the glasses of sand near its base, tapping the top of one of the balls of playdough, picking up one of the glasses of sand).

To prepare for coding the gesture-speech relationship on a gesture-by-gesture basis, we divided children's verbal responses into clauses, each of which included a main verb. We then viewed the videotapes again and identified the words that coincided with each individual gesture in the gestured response. We then classified each individual gesture as either *redundant* with respect to the information in the accompanying verbal clause or *non-redundant* with respect to the information conveyed in the accompanying verbal clause.

This system allowed for four types of gestures: redundant substantive gestures, non-redundant substantive gestures, redundant indicating gestures, and non-redundant indicating gestures. Simple indicating gestures produced along with deictic expressions were always considered redundant. Non-redundant indicating gestures occurred only when more than one object was indicated within a verbal clause that referred to a single object. Examples are presented in Table 3.

Note that it is not the case that any response that includes a non-redundant gesture is automatically a *mismatch* at the response level. It is possible for an individual gesture to be non-redundant with respect to the

TABLE 2  
Response-level coding of gesture-speech relationship

Task	Speech	Gesture
Sand into short, fat glass	“This one’s bigger; this one’s fatter.” <i>Size of untransformed glass, width of transformed glass</i>	Indication-only Match Responses 1. RH point to base of untransformed glass 2. RH point to base of transformed glass <i>Indicate both glasses</i>
Playdough into flat pancake	“That one’s flat; that one’s rolled.” <i>Shape of both masses of playdough</i>	1. LH point to top of transformed playdough 2. LH point to top of untransformed playdough <i>Indicate both masses of playdough</i>
Sand into short, fat glass	“This is just, um really fat.” <i>Width of transformed glass</i>	Substantive Match Responses 1. Both hands in “C” handshapes at sides of transformed glass <i>Width of transformed glass</i>
Playdough into flat pancake	“This one’s round and this one’s flat.” <i>Shape of both masses of playdough</i>	1. RH grasp handshape over untransformed playdough 2. LH flat palm over transformed playdough <i>Shape of both masses of playdough</i>
Sand into short, fat glass	“That one has much more because it’s round.” <i>Amount of sand and shape of transformed glass</i>	Mismatch Responses 1. RH “C” handshape held around untransformed glass 2. LH point to level of sand in transformed glass 3. LH point at top edge of transformed glass, traces part of rim <i>Width of untransformed glass, level of sand in transformed glass, height of transformed glass</i>
Playdough into sausage	“Because it’s this bigger.” <i>Size of transformed playdough</i>	1. Both palms held at ends of transformed playdough (sausage) <i>Length of transformed playdough</i>

TABLE 3

Individual-gesture-level coding of gesture-speech relationship (parentheses in the speech transcript indicate an accompanying gesture)

<i>Task</i>	<i>Speech (separated into clauses)</i>	<i>Gesture</i>	<i>Type of gesture</i>	<i>Relationship</i>
Playdough into flat pancake	[This one's fatter] and [this one's] [like a pancake.]	1. Both hands in 'C' handshapes around untransformed playdough ( <i>width</i> ) 1. RH palm over transformed playdough ( <i>flat shape</i> ) 2. RH point on top of transformed playdough, taps ( <i>indication</i> )	1. Substantive 1. Substantive 2. Indication	1. Redundant 1. Redundant 2. Redundant
Playdough into two small balls	[Cause] [these two are smaller.]	1. RH point towards untransformed playdough ( <i>indication</i> ) 2. RH spread on top of transformed playdough, slight grasp ( <i>indication</i> )	1. Indication 2. Indication	1. Non-redundant 2. Redundant
Sand into short, fat glass	[Cause it's bigger] [than that one.]	1. RH point to level of sand in transformed glass ( <i>level</i> ) 2. RH point to level of sand in untransformed glass ( <i>level</i> )	1. Substantive 2. Substantive	1. Non-redundant 2. Non-redundant
Sand into tall, thin glass	[Cause this one is more taller] and [this one] is more rounder.	1. RH point moves from base to top of transformed glass $3 \times$ ( <i>height</i> ) 1. RH point toward untransformed glass, moves left to right at sand level ( <i>level</i> )	1. Substantive 1. Substantive	1. Redundant 1. Non-redundant

accompanying clause, but for the information expressed in that gesture to appear elsewhere in the response as a whole (i.e., in a different clause). Thus, even responses classified as gesture-speech matches at the response level could (and often did) contain individual gestures classified as non-redundant.

## Reliability

The reliability of coding procedures was established by having a second trained coder evaluate a subset of the data ( $N = 24$  responses). Agreement for coding the content of spoken explanations was 95%. Reliability for coding gestures was established at multiple levels. Agreement was 91% for identifying individual gestures, 91% for classifying gestures as substantive gestures or simple indications, and 93% for assigning meanings to individual gestures. At the individual gesture level, agreement was 90% for classifying gestures as redundant or non-redundant. At the response level, agreement was 100% for classifying responses as substantive match, indication-only match, or mismatch.

## RESULTS

As described above, our analysis focused on 106 responses (53 pairs), drawn from 14 different children. In each pair, both responses focused on dimensions of the task objects, and both responses included gesture.

### Comparability of lexical access

Our first goal was to establish that lexical access was comparable across tasks. Both the Information Packaging Hypothesis and the Lexical Retrieval Hypothesis hold that spatial words are relevant for gesture generation. Thus, we examined whether the use of spatial words was comparable across tasks in the sample of responses that we identified for analysis. We first classified all of the words produced by the participants into spatial words (dimensional adjectives, nouns referring to concrete objects, verbs of motion and shape change, spatial prepositions and adverbs) and non-spatial words. We then examined the density of spatial words in participants' speech. As seen in Table 4, the proportion of spatial words used in the two tasks and the rate of spatial words per clause were comparable across tasks.

We also compared the accessibility of the spatial words used in the two tasks. The accessibility of a given word can be estimated by the frequency

TABLE 4  
Results of speech analyses

<i>Measure</i>	<i>Mean for description task</i>	<i>Mean for explanation task</i>	<i>SE of mean of paired differences</i>	<i>Degrees of freedom</i>	<i>Paired t value</i>	<i>p value (two-tailed)</i>
Proportion of spatial words	0.23	0.19	0.024	13	1.50	0.16
Spatial words per clause	0.98	0.96	0.10	13	1.15	0.88
Corpus frequency of spatial words <sup>a</sup>	4961	5143	3140	13	0.06	0.96

<sup>a</sup>The mean corpus frequency of 4961 means that, on average, the words appeared 4961 times per million words in the spoken corpora analysed by CELEX.

with which the word appears in large corpora (Levelt, Roelofs, & Meyer, 1999).<sup>2</sup> To compare the accessibility of the words used in the two conditions, we compared the spoken corpus frequency of the spatial words used in each condition, as listed in the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). As seen in Table 4, the corpus frequencies of the spatial words used in the two conditions were very similar.

Not only were the two conditions similar in terms of the corpus frequencies of the words used, but they were also similar in terms of the particular words themselves, and the frequency rank order of words within each condition. There were a total of 40 different spatial words used in the description task and 35 different spatial words used in the explanation task, and 23 of these words were identical across conditions. Of the top seven spatial words used in each task (accounting for over half of all spatial words in each task), five were identical across conditions. All of the spatial words produced in each condition, with counts aggregated across participants, are presented in Appendix A, and all of the non-spatial words are listed in Appendix B.

## Gesture analysis

Thus far, we have provided evidence that the explanation and description tasks were similar in terms of the spoken responses that children provided for them. We turn next to an analysis of gesture production in the two tasks. As we argued above, the two tasks differ in terms of the process of information packaging. If gesture is involved only in the generation of the

<sup>2</sup> According to Levelt et al. (1999), the more frequent the word, the faster the retrieval of "lexemes", the phonological information of the word. Note that Butterworth and Hadar (1989) and Krauss et al. (in press) claim that gesture helps the retrieval of lexemes.

surface forms of utterances, then gesture production in the two tasks should be comparable. However, if gesture is also involved in conceptual planning, then gesture production in the two tasks should differ.

Given the need for deictic references in both tasks, we did not expect the overall rate of gestures per clause to differ across the tasks, and indeed, it did not (description,  $M = 1.23$ ,  $SE = 0.05$ ; explanation,  $M = 1.35$ ,  $SE = 0.08$ ), paired  $t = 1.23$ ,  $p = .24$ . In addition, the mean number of clauses per response also did not differ across tasks (description,  $M = 2.23$ ,  $SE = 0.17$ ; explanation,  $M = 1.91$ ,  $SE = 0.20$ ), paired  $t = 1.14$ ,  $p = .27$ .

Recall that, under the Information Packaging Hypothesis, children should produce more gestures that include a representational component in the explanation task than in the description task. Further, children should produce more non-redundant gestures in the explanation task than in the description task. We tested these predictions both at the level of individual responses, and at the level of individual gestures.

For the response-level analyses, we calculated the proportion of responses that included at least one substantive gesture, and the proportion of responses that were classified as gesture-speech mismatches according to Church and Goldin-Meadow's (1986) system. As seen in Table 5, participants produced more responses that included substantive gestures and more gesture-speech mismatch responses in the explanation task than in the description task. Thus, both of the predictions of the Information Packaging Hypothesis were supported in the response-level analyses.

For the individual-gesture-level analyses, we assessed the proportion of substantive gestures, and the proportion of non-redundant gestures. As

TABLE 5  
Results of gesture analyses

<i>Measure</i>	<i>Mean for description task</i>	<i>Mean for explanation task</i>	<i>SE of mean of paired differences</i>	<i>Degrees of freedom</i>	<i>Paired t value</i>	<i>p value (one-tailed)</i>
Response-level analyses						
Proportion of responses with one or more substantive gestures	0.59	0.69	0.06	13	1.78	0.05
Proportion of gesture-speech mismatch responses	0.26	0.41	0.08	13	1.77	0.05
Individual gesture analyses						
Proportion of substantive gestures	0.42	0.52	0.03	13	1.53	0.08
Proportion of non-redundant gestures	0.18	0.27	0.02	13	2.14	0.03

seen in Table 5, participants produced more substantive gestures and more non-redundant gestures in the explanation task than in the description task. Thus, as in the response-level analysis, both of the predictions of the Information Packaging Hypothesis were supported in the individual-gesture-level analysis.

## DISCUSSION

The purpose of this study was to compare two accounts of the role of gesture in speaking. We sought to establish whether gesture is involved in the conceptual planning of the message, or whether it is involved only in generating the surface forms of utterances. To accomplish this goal, we developed two tasks that elicited similar utterances, but made different demands on the conceptual “packaging” of spatial information. As we argued above, conceptualisation for the explanation task was more complex and more constrained than conceptualisation for the description task.

As expected, the surface forms of utterances that children produced across the two tasks were comparable. However, children produced more substantive gestures and more non-redundant gestures in the explanation task than in the description task. These results support the view that spontaneous gestures are involved in the conceptual planning of utterances. Thus, these results are consistent with the Information Packaging Hypothesis, and do not support the Lexical Retrieval Hypothesis. Our results leave open the possibility that gesture may be involved in lexical retrieval, but they make clear that lexical retrieval is not the *only* point in the process of speaking in which gesture is involved.

How do the current findings square with other evidence reported in the literature in support of the Lexical Retrieval Hypothesis? Kita (in press) has argued that all of the evidence reported in the literature in support of the Lexical Retrieval Hypothesis can also be explained by the Information Packaging Hypothesis. The basic problem is that, in most studies, difficulty in lexical access and difficulty in conceptualisation are confounded. The present study is the first to hold lexical access constant while varying conceptualisation. As described by Kita (in press) and summarised below, three lines of putative empirical support for the Lexical Retrieval Hypothesis have been put forward in the literature. However, in all cases, the findings are also compatible with the Information Packaging Hypothesis.

First, speakers produce more substantive gestures when restrictions are imposed on their speech (e.g., avoid all words containing the letter “c”) (Rauscher et al., 1996). However, such restrictions not only make lexical access more difficult, but also require the speaker to explore alternative



ways of packaging the information to be conveyed. Under the Information Packaging Hypothesis, this is precisely the situation in which more substantive gestures should be produced. Thus, this finding is compatible with the Information Packaging Hypothesis.

Second, prohibiting gestures leads to slower speech and more dysfluencies in phrases with spatial content (Rauscher et al., 1996). In particular, prohibiting gestures leads to an increase in a specific subtype of dysfluency, non-juncture filled pauses, which fall within clauses rather than at clause boundaries. Non-juncture filled pauses purportedly derive from difficulties in lexical access (Rauscher et al., 1996); however, as Kita (in press) has argued, the studies on which this claim is based also confound lexical access difficulty and conceptualisation difficulty. Under the Information Packaging Hypothesis, prohibiting gesture should lead to difficulties in conceptualising spatial information for speaking, and this conceptualisation difficulty could lead to decreased speech rate and increased dysfluencies (including non-juncture filled pauses). Thus, these findings are also compatible with the Information Packaging Hypothesis.

Finally, the asynchrony between the onset of a gesture and its lexical affiliate is greater for less familiar lexical items (Morrel-Samuels & Krauss, 1992). However, familiarity may be correlated, not only with the accessibility of the lexical items, but also with the ease of conceptualising the information to be expressed. For less familiar information, speakers may need to explore alternative ways to package the information, so the onset of speech may be delayed. Thus, this finding is also compatible with the Information Packaging Hypothesis.

In brief, all of the putative evidence for the Lexical Retrieval Hypothesis in the literature is also consistent with the Information Packaging Hypothesis. However, the results of the current study can only be explained by the Information Packaging Hypothesis. Hence, the Information Packaging Hypothesis is to be preferred.

Of course, given the present results, we cannot make *causal* claims about the facilitatory role of gesture in the conceptual planning of utterances, because we did not experimentally manipulate gesture. Further, we acknowledge that the discourse demands of the two tasks differ, and that this may have contributed to the observed differences in the nature of the gestures produced across tasks. Nevertheless, our findings strongly suggest that substantive gestures are generated when spatial information is packaged into an informational unit to be verbalised. The findings are compatible with the view that gesture plays a facilitatory role in this process, and they converge with other evidence that suggests that gestures can play a role in thinking and memory (Alibali & DiRusso, 1999; Rime, Shiaratura, Hupet, & Ghysseleux, 1984). We are currently conducting a set of experiments that will provide causal evidence about whether gesture

plays a role in conceptualisation. In these experiments, we manipulate whether or not speakers are able to gesture, and we explore whether conceptualisation differs as a function of gesture production.

The present findings also raise the question of *how* gesture facilitates conceptualisation. What is the mechanism by which gesture influences thinking? The Information Packaging Hypothesis holds that gesture helps speakers to explore alternative ways of organising a perceptual array, and thus helps speakers to break down a perceptual array into verbalisable units (Kita, in press). In this way, gesture helps speakers to re-represent perceptual or motor knowledge in verbal form. We further suggest that gesture may also influence speakers' mental representations of situations (e.g., tasks or events) by focusing their attention on particular features of the situations (see Alibali & DiRusso, 1999). We are testing these ideas in our current research.

In sum, our data indicate that the role of gesture in speech production goes beyond lexical retrieval. We observed differences in gesture production across two tasks as a function of differences in the conceptualisation required by the tasks. Based on the Information Packaging Hypothesis, we derived specific predictions about how gesture production should differ across tasks, and these predictions were supported. We claim that the action of gesturing helps speakers to organise spatial information for verbalisation, and in this way, gesture plays a role in conceptualising the message to be verbalised. It follows from this view that gesture may play a role, not only in speech production, but also in cognitive activity more generally. Thus, we believe that any theory of human performance will not be complete without an understanding of the role of gesture in cognitive activity.

Manuscript received September 1999

Revised manuscript received April 2000

## REFERENCES

- Alibali, M.W., & DiRusso, A.A. (1999). The function of gesture in learning to count: More than keeping track. *Cognitive Development*, 14, 37–56.
- Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The Celex lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, 96, 168–174.
- Church, R.B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23, 43–71.
- Clark, H.H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- de Ruiter, J.-P. (1998). *Gesture and speech production*. Nijmegen, The Netherlands: MPI Series in Psycholinguistics.

- Engle, R.A. (1998). Not channels but composite signals: Speech, gesture, diagrams, and object demonstrations are integrated in multi-modal explanations. In M.A. Gernsbacher & S.J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 321–326). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, 27, 175–200.
- Kita, S. (in press). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action*, pp. 162–185. Cambridge, UK: Cambridge University Press.
- Krauss, R., Chen, Y., & Gottesman, R. (in press). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture: Window into thought and action*, pp. 261–283. Cambridge, UK: Cambridge University Press.
- Krauss, R.M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, 7, 54–60.
- Krauss, R.M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology*, 28, 389–450.
- Levitt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Morrel-Samuels, P., & Krauss, R. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18, 615–622.
- Piaget, J. (1967). *Six psychological studies*. New York: Random House.
- Rauscher, F.H., Krauss, R.M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7, 226–231.
- Rime, B., Shiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, 8, 311–325.
- Siegler, R.S. (1995). How does change occur? A microgenetic study of number conservation. *Cognitive Psychology*, 28, 225–273.

## APPENDIX A: Spatial words used in the description and explanation tasks

Counts are aggregated over participants and over inflectional variants (e.g., big, bigger, biggest). Capital letters indicate that the count for a lemma is an aggregation of counts from more than one inflectional variant.

<i>Description task</i>			<i>Explanation task</i>				
<i>Rank</i>	<i>Lemma</i>	<i>Count</i>	<i>Rank</i>	<i>Lemma</i>	<i>Count</i>		
1	1	BIG	19	1	1	BIG	20
2	2	SMALL	14	2	2	SMALL	8
3	3	FLAT	11	3	2	TALL	8
4	3	more	11	4	3	FLAT	7
5	3	ROUND	11	5	4	ball	7
6	6	LITTLE	7	6	4	fat	5
7	7	ball	6	7	4	ROLL	5
8	8	less	5	8	8	glass	4
9	8	LONG	5	9	8	in	4
10	8	TALL	5	10	8	LITTLE	4
11	11	circle	4	11	8	LONG	4
12	11	lower	4	12	8	pancake	4
13	11	sand	4	13	13	less	3
14	14	glass	3	14	13	more	3
15	14	in	3	15	13	out	3
16	14	much	3	16	13	round	3
17	17	fat	2	17	13	together	3
18	17	grown	2	18	18	get	2
19	17	hotdog	2	19	18	PIN	2
20	17	up	2	20	18	skinnier	2
21	21	bowl	1	21	18	tinsy	2
22	21	drink	1	22	18	bowl	1
23	21	from	1	23	23	connected	1
24	21	fuller	1	24	23	fill	1
25	21	higher	1	25	23	from	1
26	21	line	1	26	23	much	1
27	21	lot	1	27	23	outside	1
28	21	out	1	28	23	playdough	1
29	21	pancake	1	29	23	put	1
30	21	playdough	1	30	23	shorter	1
31	21	rolled	1	31	23	smashed	1
32	21	shorter	1	32	23	teensy	1
33	21	skinny	1	33	23	teeny	1
34	21	snake	1	34	23	thin	1
35	21	strawberry	1	35	23	wider	1
36	21	teeny	1				
37	21	thin	1				
38	21	tiny	1				
39	21	tower	1				
40	21	wheels	1				
		Total	143			Total	117

## APPENDIX B: Non-spatial words used in the experiment

This list is for all participants and both conditions. Inflectional variants (e.g., look, looks) are collapsed, and are indicated in capital letters (e.g., LOOK).

a	not
again	of
all	ONE
also	only
amount	really
an	same
and	see
as	size
BE	than
because	that
bit	the
both	them
but	there
DO	these
for	they
HAVE	thing
it	this
its	those
just	to
kind	two
like	usually
LOOK	very
needs	well
nice	you
no	