Modelling Learning Using Evidence from Speech and Gesture

Martha Wagner Alibali and Susan Goldin-Meadow
University of Chicago, Department of Psychology
5848 South University Avenue
Chicago, Illinois 60637

Abstract

Speech and gesture provide two different access routes to a learner's mental representation of a problem. We examined the gestures and speech produced by children learning the concept of mathematical equivalence, and found that children on the verge of acquiring the concept tended to express information in gesture which they did not express in speech. We explored what the production of such gesture-speech mismatches implies for models of concept learning. Two models of a mechanism that produces gesture-speech mismatches were tested against data from children learning the concept of mathematical equivalence. The model which best fit the data suggests that gesture and speech draw upon a single set of representations, some of which are accessible to both gesture and speech, and some of which are accessible to gesture but not speech. Thus, gesture and speech form an integrated system in the sense that they do not draw upon two distinct sets of representations. The model implies that when new representations are acquired, they are first accessible only to gesture. Over time, they are then recoded into speech.

Introduction

Learning, in both adults and children, involves moving from a less adequate to a more adequate understanding of a concept. Characterizing the process that bridges these states is crucial to understanding learning. Unfortunately, in many studies of learning in both adults and children, performance, procedures, and mental representations are described before and after learning, while little attention is paid to the transition between states (see Glaser & Bassok, 1989, for discussion).

This absence of focus on transition may be due to the difficulty inherent in studying the short-lived transitional state. What is needed is a technique for identifying when learners are on the verge of change. In previous work, we have suggested that the mismatch between the thoughts a learner expresses in speech and in gesture serves as a signal that the learner is in a state of transition. Further, we have shown that children who produce gesture-speech mismatches on a task are particularly ready to benefit from instruction in that task (Church and Goldin-Meadow, 1986; Perry, Church and Goldin-Meadow, 1988; Goldin-Meadow, Alibali & Church, 1993). For example, a child who said, "the glass is tall", while producing a gesture indicating the width of a glass on a conservation task, was more likely to benefit from instruction in that task than a child who said, "the glass is tall", while producing a gesture indicating the height of the glass. Gesture-speech mismatch thus appears to be a marker which can be used to distinguish those who are on the threshold of learning from those who are not. Moreover, and more importantly, this marker provides substantive information about the cognitive processes that characterize transitions in learning. (For further discussion of the role of gesture-speech mismatch in transition, see Alibali and Goldin-Meadow, 1993, and Goldin-Meadow, Nusbaum, Garber and Church, 1993.)

In many studies of learning, a subject is asked to solve a problem and to explain how he or she reached the solution. The subject describes a procedure for arriving at the solution, and from this procedure, inferences about that subject's representation of the problem can be made. By observing both gesture and speech, there are two different access routes to the subject's representation, one through the procedure articulated in speech and a second through the procedure described in gesture. When gesture and speech match, the two access routes provide evidence for the same representation. However, when gesture and speech fail to match, that is, when they mismatch, the two routes provide evidence for two different representations, one accessed by gesture and a second accessed by speech. Thus, learners whose gesture and speech mismatch appear to have two different representations of the same problem.

Representations accessible to gesture and to speech

The existence of such gesture-speech mismatches in learners' problem explanations has led us to question whether all of a learner's representations of a problem must be accessible to both gesture and speech, or whether some representations may be accessible to only one modality. The definition of mismatch does not entail that representations can be accessed by one modality and not the other, but only that different representations are activated in gesture and speech during a single explanation. To establish whether learners can have representations which are accessible to one modality only, one must examine a subject's entire set of explanations to determine that subject's repertoire of procedures, and investigate whether procedures appear in one or both modalities across that repertoire.

We examined the repertoires of procedures that children produced before they learned the concept of mathematical equivalence (the idea that the two sides of an equation represent the same quantity). We found that many children did indeed have procedures that they produced in one modality and not the other (Goldin-Meadow, et al., 1993). Furthermore, the mean number of procedures found in gesture but not speech was larger than the mean number found in speech but not gesture. Thus, children had a relatively large number of procedures which they expressed in only one modality, primarily only in gesture.

Tests of two models of the generation of mismatches

A model that assumes that gesture and speech are independent systems

If a representation is accessible to only one modality, whenever a learner attempts to articulate a procedure based on that representation, that learner will not be able to produce the same procedure in both gesture and in speech. What might this imply about the mechanism by which gesture-speech mismatches are generated? One possible mechanism rests on the assumption that each learner has two distinct sets of representations, one set accessible to gesture and a second set accessible to speech. When faced with a problem, the learner samples two representations of the problem: one verbal and one gestural. On the basis of the verbal representation, the learner expresses a procedure for solution in speech, and on the basis of the gestural representation, the learner expresses a procedure for solution in gesture. We might hypothesize further that the two representations are sampled independently. That is, when asked to explain a problem, a learner samples a representation accessible to gesture and independently samples a representation accessible to speech.

According to this model, a learner will produce a gesture-speech match by sampling a representation from the set of representations accessible to gesture and, by chance, independently sampling that same representation from the set accessible to speech. A learner will produce a gesture-speech mismatch by sampling a representation from the set of representations accessible to gesture, and independently sampling a different representation from the set accessible to speech.

If this Independent Model is correct, the probability of producing a gesture-speech match in any given problem explanation should be equal to the probability of sampling a particular representation from the set of representations accessible to speech, times the probability of sampling that same representation from the set accessible to gesture. We evaluated this model with respect to our data on children acquiring the concept of mathematical equivalence. At each of three assessment points, children solved and explained six equivalence problems. Each explanation was coded as a gesture-speech match or mismatch. Children varied from 0 to 6 in the number of gesture-speech matches they produced at each of three assessment points. We classified children as Mismatchers if they produced three or fewer matches , and as Matchers if they produced four or more matches (out of six). We then calculated the number of gesture-speech matches each child would be expected to produce at each assessment point, assuming that the child samples representations accessible to gesture independently of sampling representations accessible to speech. We classified children as Matchers or Mismatchers based on the predictions of the model, and compared the predicted distribution of Matchers and Mismatchers to the observed distribution.

This model fit the data quite poorly. As seen in Figure 1, the model did not make similar predictions for all three assessment points. In fact, comparing the distributions predicted by the Independent Model to the observed distributions, we found that the model predicted a distribution of Matchers and Mismatchers which differed significantly from that actually observed at two of the three assessment points (for Assessment I, $x^2(1)=8.74$, $p<0.005$; for Assessment II, $x^2(1)=29.53$, $p<0.001$; and for Assessment III, $x^2(1)=0.12$, $p>0.50$).

Furthermore, the Independent Model did not accurately model the distribution of the precise numbers of matches the children produced. We compared the distribution of the number of matches predicted by the Independent Model to the distribution of the number of matches that children actually produced. The Independent Model predicted a distribution which differed significantly from the observed distribution at two of the three assessment points (for Assessment I, D=0.237, $p<0.05$; for Assessment II, D=0.357, $p<0.01$; and for Assessment III, D=0.139, $p>0.20$; Kolmogorov-Smirnov One-Sample Test, Siegel, 1956). Thus, at least one of the two assumptions upon which this model is based, namely, that a learner samples two representations, and that the learner samples them independently, is not tenable.

A model that assumes that gesture and speech form an integrated system

As an alternative, we suggest that gesture and speech draw upon a single set of representations, some of which are accessible to both gesture and speech, and some of which are accessible to gesture but not speech. When faced with a problem, a learner samples a single representation of the problem and, on the basis of that representation, describes a procedure for solving the problem. If the learner samples a representation which is accessible to both gesture and speech, the learner will express the same procedure in both modalities, thus producing a gesture-speech match.

If, however, the learner samples a representation which is accessible to gesture but not to speech, he or she will be able to describe the procedure in gesture but will be unable to express the same procedure in speech. In this case, the learner has two options. As the first option, the learner may select another representation of the problem to articulate in speech. In this case, the learner will produce a gesture-speech mismatch. As a second option, the learner may attempt to recode or redescribe that representation into speech. If the learner succeeds, the procedure based on that representation will then be accessible to both gesture and speech, and the learner will produce a gesture-speech match. If the learner does not succeed, he or she is likely to produce speech which is ambiguous or uninterpretable (see also Graham and Perry, in press, and Siegler and Jenkins, 1989, for arguments that learners on the brink of change often produce vague or inexplicit spoken explanations). Such ambiguous or uninterpretable speech, accompanied by a gesture which conveys a clear problem solving procedure, also constitutes a gesture-speech mismatch.

Thus, according to this model, if a representation which is accessible only to gesture is the first representation sampled, the learner will produce a gesture-speech match only if he or she can successfully recode that representation into speech. Otherwise, the learner will produce a gesture-speech mismatch. If a representation which is accessible to both gesture and speech is the first representation sampled, the learner will produce a gesture-speech match.

If the Integrated Model is correct, the probability of producing a gesture-speech match on any given problem should be equal to the probability that a representation which is accessible to both gesture and speech will be sampled (either one that was previously part of the learner's repertoire, or one that was spontaneously recoded and is now accessible to both modalities). To evaluate this model with respect to our data on children learning mathematical equivalence, we recalculated the number of gesture-speech matches each child would be expected to produce at each assessment point assuming that gesture and speech form an integrated system. We then classified children as Matchers or Mismatchers based on the predictions of the model, and compared the predicted distributions of Matchers and Mismatchers to the observed distributions.

As seen in Figure 1, the Integrated Model made similar predictions for all of the three assessment points. Comparing the distributions predicted by the Integrated Model to the observed distributions, we found that the model predicted a distribution of Matchers and Mismatchers which did not differ significantly from that actually observed at each of the three assessment points (for Assessment I, $x^2(1)=0.44$, $p>0.50$; for Assessment II, $x^2(1)=1.78$, $p>0.15$; and for Assessment III, $x^2(1)=0$, $p=1.0$). Thus, the Integrated Model provided a better fit to the data than did the Independent Model.

Furthermore, the Integrated Model also accurately modelled the distribution of the precise numbers of matches the children produced. At each of the three assessment points, the Integrated Model predicted a distribution of matches which did not differ significantly from the observed distribution (for Assessment I, D=0.105, $p>0.20$; for Assessment II, D=0.143, $p>0.20$; for Assessment III, D=0.167, $p>0.20$; Kolmogorov-Smirnov One-Sample Test, Siegel, 1956).

Thus, the model that best fit the data assumes that the child samples representations which are then encoded into gesture and/or speech. According to the model, a child has a single set of representations, some of which are accessible to both gesture and speech, and some of which are accessible only to gesture. Further, the model assumes that when a representation which is accessible to both gesture and speech is the first representation sampled, both gesture and speech will be activated. In this sense, this model is consistent with McNeill's (1992) description of gesture and speech as an integrated system.

Discussion

Based on these findings, we propose the following description of the steps a learner takes in acquiring the concept of mathematical equivalence. The learner begins the acquisition process with incorrect representations of the concept, most of which are accessible to both gesture and speech. The learner then acquires correct representations that are accessible only to gesture and not to speech. At this point, the learner is unable to verbally express those representations which are accessible only to gesture, and therefore, he or she is likely to produce gesture-speech mismatches. At this moment, the learner is in a

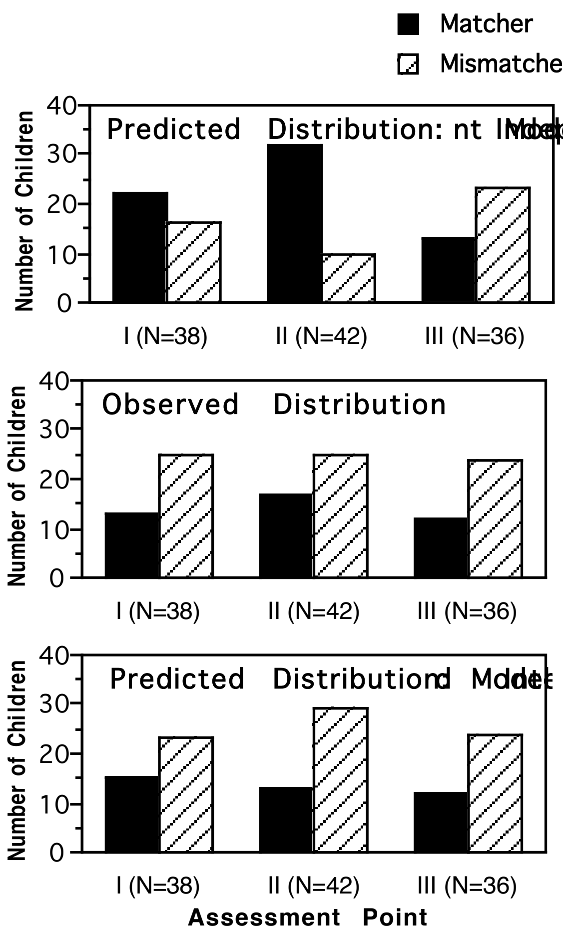

Figure 1. Distributions of Matchers and Mismatchers (1) predicted by the Independent Model, (2) observed, and (3) predicted by the Integrated Model. At each assessment point, children who did not produce any procedures in gesture were excluded from the analyses.

transitional state with respect to this concept and is most open to instruction (cf. Church and Goldin-Meadow, 1986, Perry et al., 1988). Finally, the learner develops a verbal code for the correct representations that were once accessible only to gesture, returning once again to a state in which most of his or her representations are accessible to both gesture and speech. This time, however, the representations are correct.

Why might a representation be accessible to gesture but not to speech? McNeill (1992) has argued that gesture and speech are two aspects of a single process. The two modalities are correlated in meaning but do not always reveal the same meaning. According to McNeill, gesture reflects a global, synthetic image which is idiosyncratic and constructed at the moment of speaking. In contrast, speech reflects a linear-segmented, hierarchical linguistic structure, which draws on a conventional, socially-constituted grammar and lexicon.

Following McNeill, we suggest that gesture is a vehicle for conveying ideas which are based on images. Gesture offers learners a vehicle that is distinctly different from speech for expressing their understanding of a problem. Our data suggest that, for certain problems and at certain times in the learning process, gesture may be better suited to capturing a learner's understanding of a problem than is speech.

References

Alibali, M. W., & Goldin-Meadow, S. In press. Gesture-Speech Mismatch and Mechanisms of Learning: What the Hands Reveal about a Child's State of Mind. *Cognitive Psychology*.

Church, R. B., & Goldin-Meadow, S. 1986. The Mismatch Between Gesture and Speech as an Index of Transitional Knowledge. *Cognition*, 23:43-71.

Glaser, R. & Bassok, M. 1989. Learning Theory and the Study of Instruction. *Annual Review of Psychology*, 4:631-666.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. 1993. Transitions in Concept Acquisition: Using the Hand to Read the Mind. *Psychological Review*, 100(2):279-297.

Goldin-Meadow, S., Nusbaum, H., Garber, P, & Church, R. B. 1993. Transitions in Learning: Evidence for Simultaneously Activated Hypotheses. *Journal of Experimental Psychology: Human Perception and Performance*, 19: 1-16.

Graham, T., & Perry, M. In press. Indexing Transitional Knowledge. *Developmental Psychology*.

McNeill, D. 1992. *Hand and Mind*. Chicago: University of Chicago Press.

Perry, M., Church, R. B., & Goldin-Meadow, S. 1988. Transitional Knowledge in the

Acquisition of Concepts. *Cognitive Development*, 3:359-400.

Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Siegler, R. S. & Jenkins, E. 1989. *How children discover new strategies*. Hillsdale, N.J.: Lawrence Erlbaum Associates.