Image Schemas and Gesture[*]

Alan Cienki

1.   Introduction

Beginning with Johnson (1987) and Lakoff (1987), the vast majority of research on image schemas has involved the analysis of linguistic data. Semantic analyses in terms of image schemas have received particular attention, especially as they relate to metaphoric use of language that is argued to map from image schemas as source domains. However, if we consider Johnson's characterization of image schemas in the Preface to his 1987 work, it was broadly comprehensive:

> An image schema is a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience. (Johnson 1987: xiv).

How do image schemas relate to other aspects of our experience? The question is enormous, and with the present study, I propose to move modestly in scope by looking not only at linguistic utterances, but also at co-verbal behavior, namely manual gesture with speech. Though the study of gesture dates back at least to classical antiquity,[1] in the twentieth century in particular it underwent a "recession and return" (Kendon 2004: chapter 5). Gesture studies, as a field of inquiry, has now become the subject of regular international conferences, a journal entitled *Gesture*, and the basis for the formation of the International Society for Gesture Studies. Yet the

[1] See Müller (1998) and Kendon (2004) for historical overviews.

bodies of research on image schemas and gesture studies have barely made any connection with each other.[2]

Though language and gesture have been argued to constitute aspects of a larger conceptual/communicative system (McNeill 1985, 1992), there are important differences in how the two modalities are capable of expressing ideas. Language (in terms of sound or written symbolic form) is a "linear" code that is digital in its ontology, while gesture, with its multi-dimensional forms, is analog in nature (McNeill 1992: 11). Given that image schemas, too, are proposed to exist "in a continuous, analog fashion in our understanding" (Johnson 1987: 23), it seems natural that gestures might embody image schemas. Speech and gesture also play different roles in the development of our thoughts in real time. Gesture, it has been argued, can both reflect and influence the imagery content of an idea unit as it is being "unpacked" during speech (Kita 2000; McNeill and Duncan 2000). This also suggests a potential connection between gestures and the imagery inherent in image schemas.

A number of other factors suggest that gesture should provide fertile ground for research on image schemas. One concerns the multifaceted character of gestures and the gestalt nature of image schemas. For example, image schemas constitute patterns which can be thought of in either a static or dynamic fashion, realized as an entity or a process (Cienki 1997: 6-7). PATH can be understood as the linear motion of something or the static trace of that motion (or a potential trajectory); CONTAINER, though normally experienced as an entity, can be construed through the continued motion of an object in a cyclical path; and similar options hold for most other image schemas. Manual gestures also consist of physical form and motion: a form of the hand and forearm, and a motion in the main gesture stroke phase, optionally followed by a hold in a static position. The motion of the gesture may itself be intended to describe or outline a static form in the air. As Müller (1998: 114-126) explains, in gesturing, the hand can act (as it would in carrying out an activity with object), it can model or sculpt something in the air, it can draw an outline (trace), or it can itself represent a form as if a sculpture. We know from sign language research (Taub 2001) that the hands can be used to express concepts in a richly iconic way, far more than is possible with the oral means by which spoken language is produced. One hypothesis, then, is that image schemas may not only serve

---

[2] For some initial work bridging this gap, see Cienki (1997, 2002) and Calbris (2003).

as structuring elements of the semantics of verbal expressions; they may also constitute structuring elements of gestures, informing the iconic potential of manual expression.

An alternative approach, however, would be to focus on the fact that gesturing varies considerably across individuals, even within a given cultural group. McNeill (1992: 1), for example, argues that co-speech gestures are "free and reveal the idiosyncratic imagery of thought." He adds that they are shaped by the individual speaker's own meaning rather than by any shared conventions: "in no sense are they elements of a fixed repertoire" (*ibid.*). Even though "[t]he gestures of different speakers can present the same meanings[, they] do so in quite different forms" (ibid.: 22). On this logic, an alternative hypothesis is that gestures may be too variable in form to allow them to be related to patterns such as image schemas.

Below I present the findings of an experiment which assesses the degree to which gestures can be interpreted as image-schematic in form. The first part describes how the gesture stimuli for the experiment were selected and prepared. Then the experiment itself involves four different conditions. Participants in each were given a set of image schemas as descriptors and asked to use one of them or the word "other" to characterize either (a) a series of gestures without sound, (b) the same gestures with the accompanying speech, (c) the accompanying speech (phrases) only, in audio and written transcript form, or (d) the phrases in written transcript form only. The intent was to find out whether image schemas could reliably be used as descriptors for the gestures, and to assess what role the accompanying speech plays in trying to interpret the gestures in terms of image schemas. A secondary factor considered is whether gestures of different functional types, and the utterances that accompany them, are characterized differently in terms of image-schema descriptors or not.

2.   Experiment

2.1   Preparation of the materials

The source of the video-recorded gestures for the experiment was a set of six dyadic conversations, which are part of a larger cross-linguistic study on metaphors for 'honesty.' The conversations for that study were elicited

between pairs of friends, all speakers of American English, who were undergraduates at an American university. At the beginning of each session, the student dyad was given a set of four large cards (three with questions and one with a very short story) that prompted them to discuss how they take exams at their university. Each conversation lasted approximately 20 minutes.

Gestures occurring in the same two-minute segment of each of the six videos were analyzed (thus 12 minutes of video material in total). In each case minutes 3 to 5 were used. This provided comparable content across the tapes as the speakers were all engaged at a similar point in the procedure. Since the time period is a few minutes into the procedure, the participants had gotten warmed up and were more involved in the task, and so one could expect more natural gestural behavior.

2.1.1 Preparation step 1

The first step involved selecting which movements would be considered gestures. Gestures were coded by the author and a second trained analyst to test reliability. Our baseline was identifiably distinct effortful movements of the hands and forearms, that is, gesture strokes. We ignored "self-adjustors," such as scratching oneself. The stroke phase of a gesture may be preceded by a preparation movement to get the hand into a ready position, and/or it may be followed by a retraction movement (McNeill 1992). If so, the preparation/retraction was included with the stroke phase that it led up to or returned from, and was not counted as a separate gesture. Dividing points between gestures were determined by several criteria for judging the release of a distinct effortful movement, including return of the hand to a rest position, and/or relaxation of the hand between tense hand shapes. Previously prepared written transcripts of the speech in the videos were marked for gesture strokes by bracketing the speech (or pauses, which were also noted on the transcripts) which accompanied them.

Of the 212 movements identified by at least one coder as a gesture stroke in the 12 minutes of video, we agreed on 156 as gestures, or 74%. Through subsequent discussion of individual cases we were able to bring our agreement to 100%, resulting in a set of 189 gestures, which were the starting point for Step 2.

2.2.2 Preparation step 2

The second part involved coding the gestures according to function. For this step, an adaptation of Müller's (1998) functional classification was used. This system was preferred because of its consistent focus on functional criteria, as opposed to some other systems which mix formal and functional criteria[3]. Müller draws on Bühler's (1982) "Organon" model of language, which describes three functions of a linguistic sign in every speech situation, namely: representation (*Darstellung*), expression (*Ausdruck*), and appeal (*Appell*). That is, according to Bühler, every use of language (a) represents entities, relations, etc.; (b) expresses the inner, subjective perspective (*Innerlichkeit*) of the speaker; and (c) is directed to (appeals to) a particular addressee. One function is normally highlighted more than the others in a given use of language, but the others remain in the background. Müller elaborates on the application of the three functions to gesture, noting that gestures can also be representational (e.g., depicting the shape of a picture frame), expressive (showing elation by clapping one's hands), and appeal to (be directed toward) an address (as with a gesture giving a blessing). Like linguistic signs, gestures serve multiple functions at once, although normally one is highlighted. For example, the manner in which one performs a primarily representational gesture can show more or less saliently "expressive" and "appeal" functions, e.g., one's affective stance toward the subject under discussion, and towards the addressee.

From this starting point, Müller (1998: 110-113) focusses on the representational use of gestures, and proposes a classification of three basic functional types: referential gestures (depicting concrete and abstract entities, relations, actions, etc.), discursive gestures (which structure the accompanying verbal utterance, for examply by marking emphasis), and

---

[3] Müller (1998: 91-103) provides a critique of the most widely used classification systems. For example, with Eckman and Friesen's (1969) classification, the category of "illustrators" includes gestures based on form (deictics), function ("batons," which mark emphasis and can take various forms), and referent ("pictographs" and others, which depict entities and movements). McNeill's (1992) four categories of gestures — beats, deictics, iconics, and metaphorics — conflate formal, functional, and semantic criteria. Thus the formally determined category of deictic gestures can be referential or discourse-structuring in function. The iconic nature of metaphoric gestures cannot be accounted for because of the distinction between iconic and metaphoric gestures, and the two categories are also based on different criteria — formal versus semantic.

performative gestures (which enact a speech act, such as dismissing an idea). The first category, referential gestures, encompasses two subcategories: concrete referential gestures (iconically depicting something physical being talked about), and abstract referential gestures (depicting an abstract idea with a physical form, and in this sense metaphoric). The basic types of the classification system are shown in Table 1.

Referential gestures		Discursive gestures	Performative gestures
/ \			
Concrete reference	Abstract reference		

Table 1: Main categories in Müller's (1998) classification of gestures by function

The multifunctionality of gestures is also relevant across the categories of Müller's classification. Therefore, to use it as a basis for coding, one must focus on interpreting each gesture's primary function. For example, a gesture in which the speaker counts off the logical points in an argument on consecutive fingers on one hand would be classified as a discursive type of gesture here, as the primary function is to assist in structuring the argument. Even though the gesture also entails abstract reference as ideas are being rendered as entities (the different fingers), in the context of sequential counting, this function is subservient to the more salient function of distinguishing different ideas presented in the discourse. However, the primary function could shift to abstract reference with an appropriate change in context. For example, the speaker could change tactic from counting, and instead dwell on one particular point by saliently moving one finger out and possibly holding it out for a longer period of time. Such a gesture highlighting reference to the idea (an abstract entity) as an object (the finger) would be categorized as abstract referential. The speech context of the gesture is therefore essential for assessing its primary function.

An earlier attempt to use the four-way classification as a coding scheme for gestures proved problematic. In a previous study (Cienki 2003) in which it was applied, Cohen's Kappa for identification of function between two coders for 105 gestures was low at .25. One important problem concerned the coding of some gestures as primarily performative or discursive in function: for example, there was difficulty in discriminating between presenting an idea (performative) or emphasizing an idea (discursive) as the primary function of some gestures. Consequently, an adaptation of Müller's

classification has been used here which distinguishes concrete referential (C) and abstract referential (A) from "other" (O) gestures, the last category encompassing the two non-referential types (performative and discursive gestures). See Table 2 for details.

   After training with the coding system, the author and a research assistant independently coded the set of 189 gestures identified in Step 1. Cohen's Kappa for our identification of gesture function was moderate at .59. Discrepancies were resolved through subsequent discussion, resulting in agreement in categorizing 28 gestures as C, 32 as A, and 129 as O.

Concrete reference [C]	Abstract reference [A]	Other [O]
• Objects (e.g., a picture frame)	• Entities (e.g., the frame-work of a theory)	• Actions (e.g., dismissing, requesting, swearing, hand clapping)
• Properties (e.g., the straight edge of a ruler)	• Properties (e.g., honesty as straight and solid)	• Emphasis (e.g., through beats)
• Behaviors and Actions (e.g., the rolling of a tire)	• Behaviors and Actions (e.g., the "rolling" de-velopment of a process)	• Structuring (e.g., with counting gestures)
• Relative location (e.g., the space behind oneself)	• Relative location and relative time (e.g., the past as behind oneself)	• Presenting (an idea or argument)

Table 2: The classification of gestures used, adapted from Müller (1998: 113)

For the experiment to follow, the stimuli were restricted to the A and O gestures (and/or the accompanying phrases, as described below). There were several reasons for this. One is that many of the C gestures in the data set were deictic (pointing) gestures. As is clear from Kita (2003), pointing gestures have unique properties and so deserve special treatment. A future study (Cienki, in preparation) will be conducted with the concrete referential gestures for comparison with the experiment reported below. Another reason that C gestures will be discussed separately is that we already know from Beattie and Shovelton (1999, 2001) something about such gestures ("iconic" as per McNeill's terminology) in relation to their ability to communicate. Abstract referential gestures, on the other hand, have (arguably) received less attention in the recent gesture literature. This is due, in part, to the research tradition of studying the gestures of speakers retelling the story of a cartoon or film they have just viewed. Such studies are necessarily biased toward concrete referential uses of gesture. A third, practical reason is that due to logistical and time limitations, it was only feasible to compare a substantive number (20) of two different types of

gestures (so, 40 in total) and their accompanying phrases in this experiment. In light of the two other considerations mentioned above, the A and O gestures were the two types chosen for this experiment

2.3 *Materials and experimental conditions*

Twenty of the A gestures and 20 of the O gestures were randomly selected for use in the experiment. The videos, which were in QuickTime format, were edited to produce individual clips of the gestures selected, each clip lasting only one to two seconds. The written transcription of the speech accompanying each gesture was also noted for each clip. The relevant transcribed utterances will be referred to as "phrases" below; most were noun phrases or verb phrases, but a few consisted of only one word.
There were four experimental conditions, as follows.

Condition 1: viewing 40 gesture clips (no sound and no transcript).
Condition 2: viewing 40 gesture clips with sound and reading transcription of the accompanying speech.
Condition 3: hearing the 40 phrases uttered on the clips and reading transcription of them (no video).
Condition 4: reading transcription of the 40 phrases uttered (no video and no sound).

The different conditions were used to distinguish differences in the interpretation of the visual cues of gesture, aural cues of prosodic features of speech, and semantic cues of the utterances themselves. Specifically, conditions 3 and 4 were used as controls to establish what was being interpreted based on gesture, and what based on the verbal utterances.[4] In each condition the "A" and "O" stimuli, be they gestures or verbal phrases, were mixed in random order, and multiple orders were used within each condition to counterbalance.

---

[4] In condition 2, the written transcription of the utterances was provided because the sound was not of high enough quality on all of the clips to discern what was being said.

2.4   Participants

Eighty Emory University undergraduate students, 20 for each condition, participated for introductory psychology course credit. All were native speakers of American English.

2.5   Procedure

For conditions 1 and 2, the video clips of the gestures were projected from a laptop computer onto a screen with a data projector. For conditions 2 and 3 (with sound), the audio of the spoken phrases was amplified with speakers. Each clip was played approximately four times. The stimuli were presented to participants in small groups, but each participant completed his/her own score sheet independently without consultation with the others.

  The scoring sheet for conditions 1 and 2 began with the following instructions:

> You will view a series of 40 video clips, each 1-2 second long. Each clip shows a person making a hand gesture. You will view each clip four times, after which you should circle one word from those given which you think best characterizes the form of the gesture. You may think some examples could be characterized by more than one of the choices given, but just pick the one that fits the best.

The instructions for condition 3 were altered appropriately to say "You will hear a series of phrases recorded from conversations," and in condition 4: "You will read a series of phrases transcribed from conversations." In both conditions 3 and 4, the task given was to circle the word which "best characterizes the phrase." In conditions 1, 2, and 3, the presentation of the video and/or audio stimuli controlled the pace of the procedure. Since condition 4 did not involve the presentation of video or audio stimuli, each stimulus phrase was written on a separate slip of paper followed by the seven descriptors, and the 40 slips of paper were stapled together as a packet. This was designed to encourage focus on the individual phrases, similar to that effected by the multiple playing of the stimulus clips in the other conditions.

  Finally, the instructions included a listing and brief explanation of the descriptors which would be provided (the names of the image schemas plus "other"). Six image schemas were used from the list of 27 presented in Johnson (1987). Since the image schemas in Johnson's list vary greatly in

character – ranging from the very general level, such as PROCESS, to the very detailed, such as RESTRAINT REMOVAL — the ones chosen were at an intermediate level of specificity. The logic here is that this would lessen the attention to degree of specificity as a factor for selecting one image schema over another, and instead leave the focus on the form/nature of the image schema itself. Also for this reason, the various kinds of forces which Johnson discusses were represented here simply as FORCE because, as he notes, they share so many properties separate from their directionality, and thus FORCE constitutes a gestalt structure (Johnson 1087: 44).

The descriptors were introduced with the wording: "Here are ways to think about the words which will be used," and each was explained and exemplified in non-technical terms, as shown in Table 3. The word "other" was accompanied by a blank to give participants the option of writing in their own choice of a descriptor, independent of the image-schema options provided. The descriptors were listed in random order for each clip/phrase. After going through the procedure, participants filled out a brief background questionnaire, and then were debriefed about the experiment.

container	A container has a boundary that separates an inside from an outside. It can hold things. We can be contained (for example, in a room), and our own bodies are containers.
cycle	A cycle begins, proceeds through a sequence of connected events, and returns to the original state to start anew. We experience cycles through time in nature and in our lives.
force	Force usually implies the exertion of physical strength in one or more directions. We can experience force in terms of compulsion, attraction, blockage, or enablement.
object	An object is a material thing which we can see and feel. We may think of an object as a discrete item.
path	A path is a route for moving from a starting point to an end point. We can follow an existing path, or make a path with our own movement.
other	If none of the words above seems appropriate, circle "other" and write in a word of your own choosing which you think best describes the gesture.

Table 3: Description of image schemas used in the experiment

2.6   Results

The responses were tallied and analyzed in the following ways. The frequency of use of the "other" category was first assessed as a measure of how strongly or weakly a given gesture clip or transcribed phrase evoked a schema. Then I examined the consistency with which the favored image schema was used as a descriptor for each gesture or phrase across participants. The amount of agreement in the choice of descriptor was then compared within and across the categories of stimulus medium (gesture versus transcribed phrase). Finally, the results from the A versus O stimuli were compared across conditions for any significant differences according to several different measures.

First, the response category labelled ~~of~~ "other" was seldom used in any of the conditions. It was never the descriptor which received the most responses for any stimulus, and in fact the mean response for this category was consistently below 1 (Mean = .93, range = 0-5 for condition 1, .55, range = 0-3 for condition 2, .76, range = 0-3 for condition 3, and .13, range = 0-1 for condition 4). Given that for each gesture viewed or phrase heard or read there were seven choices (six image schemas and an "other" category), the probability of selecting any individual descriptor given random responding would be 14% (1/7 choices). But this descriptor was consistently chosen at below chance rates, t's (39) = -11.40 for condition 1, -19.47 for condition 2, -18.88 for condition 3, and -51.59 for condition 4, p's < .001. This is an initial indication that the participants were receptive to using image schemas as possible descriptors for the gestures and phrases.

The data were then examined to ascertain how much agreement there was among participants in classifying the gestures without and with sound (conditions 1 and 2), and the phrases with and without sound (conditions 3 and 4), using image schemas as descriptors. In each condition, I calculated the number of individuals who classified the gesture or phrase into its most frequent image schema category for each item in each condition. Beginning with the gesture analyses (conditions 1 and 2), while almost all of the stimuli elicited some variable responding, the data revealed that participants selected image schemas as descriptors with more consistency than would be predicted by chance in both conditions, without and with sound (Mean = 9.5, range = 5-16 for video alone, in condition 1; Mean = 9.3, range = 5-17 for video with sound and transcript, in condition 2). Considering the categorization of the transcribed phrases, here too participants chose image schemas as descriptors more consistently than would be predicted by

chance, both with and without sound (Mean = 8.6, range = 4-14 for transcript and sound, in condition 3; Mean 8.7, range = 5-16 for transcript alone, in condition 4). The four types of stimuli elicited consistent responding at above chance rates, t's (39) = 13.02 for condition 1, 15.19 for condition 2, 14.52 for condition 3, and 14.86 for condition 4, p's < .001. On the whole, there is reliable agreement about what schema each gesture or phrase is evoking, at least within any given condition of presentation.

The results for conditions 3 and 4 may appear problematic, given that these conditions were meant to serve as controls. However, the question arises: was the same schema being chosen to characterize a gesture as was used to characterize the phrase that accompanied it? A closer look at the responses reveals that this is often not the case. For example, in one clip one of the speakers says "you can learn more that way sometimes." In conditions 3 and 4, where just the phrase was given, 11 out of 20 participants in each condition selected PATH as the descriptor. However, in each of conditions 1 and 2, in which participants viewed the gesture, 17 out of 20 chose CYCLE. This is because the speaker makes a gesture with his two index fingers extended towards each other and revolving around each other as his hands move toward himself, as if depicting a rolling ball. While this was salient to participants in conditions 1 and 2, those in conditions 3 and 4, only had the accompanying phrase to rely on, and perhaps were influenced by the use of the word "way," or the notion of learning process as metaphorical motion along a path. To compare the amount of agreement versus disagreement on the image schema chosen most frequently as the descriptor, a series of Fisher's exact tests were performed across pairs of the conditions. Given the variable factors of medium (gesture versus transcribed phrase), sound (with or without), and video (viewing the gesture or not), the only comparison yielding a significant result was between conditions 1 and 2 and conditions 3 and 4: this revealed higher agreement within media (gesture versus transcript) than between media (Fisher's p=.02).

Finally, we will consider the factor of gesture function. Was one group of gestures (A or O) easier to characterize in terms of image schemas than the other? Did the same hold true for the phrases affiliated with them? If we look at the results in condition 1 divided according A and O gestures and compare the mean number of most frequent responses for each group, a two-sample t-test shows significantly more agreement in categorizing the A gestures without sound than the O gestures, Mean = 10.5, range = 6-16 for A gestures, 8.5, range = 5-14 for O gestures, t (38) = 2.04, p=.05. A

comparison of the gestures shown with sound (condition 2) shows only marginally more agreement for the A gestures than for the O gestures, Mean = 10.0, range = 6-17 for A gestures, 8.7, range = 5-14 for O gestures, t (38) = 1.49, p=.07. Comparisons of the mean most frequent responses for the phrases that accompanied the A and O gestures (conditions 3 and 4) do not reveal any significant differences (p's > .05).

Another way to examine consistency of image-schema classifications was to determine how many image schemas were used at least once by the 20 participants per stimulus within each of the two groupings (A versus O) in each condition. A significantly lower number of image schemas used for one group would indicate more reliable agreement about image schemas as descriptors for that group. T-tests comparing the responses for the two groups in each condition only showed a significant difference in condition 2, where for the gestures with sound a significantly higher number of schemas was used for O gestures then A gestures, indicating more agreement on the characterization of the A gestures: Mean = 5.0, range = 3-7 for A gestures, Mean = 5.7, range = 4-7 for O gestures at this threshold of use at least once, t (38) = 2.36, p=.02.

An additional test considered the number of stimuli in each group for which all seven categories were selected as descriptors at least once. A lower number of such clips in a group would point to greater agreement about the image-schema descriptors for those gestures/phrases, since participants would have settled on a smaller number of appropriate image schemas for each of them. Significant differences were found only in conditions 1 and 3. Whereas seven of the O gestures without sound (condition 1) elicited responses for all seven descriptors, only one of the A gestures did so. This difference was found to be significant, p=.02, using Fisher's exact test, again indicating greater agreement in characterizing A gestures. With three of the O gestures and two of the A gestures with sound in condition 2 eliciting responses for all seven descriptors, the difference was not significant. Surprisingly, in a pattern comparable to condition 1, ten of the phrases with sound (condition 3) that had co-occurred with O gestures garnered responses for all seven descriptors, while only three of the phrases from A gestures did. This difference was also significant at Fisher's p=.02, showing greater agreement in characterizing phrases with sound that accompanied A gestures than O gestures. However, only one A phrase and one O phrase without sound were characterized using all seven categories in condition 4. In no condition were the O gestures or affiliated phrases characterized with greater agreement than the As in any of the

conditions, according to this measure. The results of the three tests comparing A and O stimuli, described above, are summarized in the Appendix.

Then a 3 x 2 (Medium [gesture versus transcript] x Sound [on or off] x Condition [Abstract versus Other]) analysis of variance (ANOVA) was performed with condition as the between-subjects factor (with the individual clips serving as the subjects) and medium and sound as within subjects variables. While there were no overall main effects, there was a marginal three-way interaction of medium by sound by condition, $F (1, 38) = 4.04$, $p=.052$. A post-hoc analysis using Tukey's honestly significant difference (HSD) found that there was a higher consensus for characterizing the A gestures without sound (from condition 1) than the A phrases (transcripts) without sound (condition 4). This was not true for the A gestures versus A phrases with sound, or for any comparisons between the O gestures and phrases.

3. Discussion

3.1 Interpreting the present results

One general conclusion we can draw from these findings is that gestures provide easily accessible manifestations of image schemas. As commonly recurring dynamic patterns of our perceptual experience and motor programs (Johnson 1987), image schemas are readily available, indeed "on hand," for recruitment as gestural forms. The ease with which many image schemas can be represented in gesture, either as static entities or dynamic processes, reinforces the motivation for the connection. For example, PATH can be indicated with one's forearm and hand outstretched, or by moving one's hand in a line and tracing a path; SURFACE can be gestured with a flat hand, held in position, open and tense, or by sweeping a flat hand through space, tracing a surface.

A second basic finding is that gestures can depict, or invoke, different schemas than speech alone can. This suggests one way in which they can make more information manifest and available to discourse participants. One case in point discussed earlier concerned the different characterizations of a gesture (index fingers rotating as if showing a rolling ball) and the co-

occuring speech ("you can learn more that way sometimes") as CYCLE and PATH, respectively. In this instance there was majority agreement among respondents (over 50%) in characterizing both types of stimuli with the different image schemas. However, in some cases, there was high agreement in characterizing the gesture, but not the phrase uttered, or vice versa. So in one clip the speaker said, "So in general on a regular test," and the highest number of responses in the transcript-only condition was 5 out of 20, and this occurred for two image schemas, CONTAINER and OBJECT. Similarly, in the transcript plus audio condition, the highest number of responses was five, and again two image schemas were selected this often this time, CONTAINER and CYCLE. In the gesture which accompanied this phrase, the speaker put both of her hands out in front of her, flat and palms down, and moved them simultaneously outwards laterally. The majority of participants in both of the gesture-viewing conditions chose SURFACE as the best characterization (16 in the gesture-only condition, and 12 in the gesture plus sound condition). The reverse pattern was sometimes found as well. In a clip in which the speaker said "to put in," there was high agreement in the transcript conditions (14 responses for CONTAINER in both conditions 3 and 4). The accompanying gesture involved a simple movement of both the speaker's loosely opened hands, beginning with her palms facing her torso and moving outward to a palm-up orientation. In the gesture plus speech condition, only six participants characterized this with CONTAINER, and in the gesture-only condition, five chose CONTAINER and five chose FORCE. These findings for A and O gestures show that the gestures can convey additional information to what is said, at least some of the time, and so the findings complement Beattie and Shovelton's (1999) conclusions about concrete referential gestures, although on the basis of a different kind of evidence.

The fact that participants were able to reliably interpret the gestures in terms of image schemas may have implications for the research on thinking for speaking. McNeill and Duncan (2000) incorporate Slobin's (1987) idea of thinking-for-speaking in terms of a growth-point model, which characterizes the development and verbal expression of an idea as the "unpacking" of a unit which combines both imagery and linguistic content. While the present experiment looked at how gestures and language could be interpreted in terms of image schemas, the findings suggest the potential relevance of image schemas (or at least these kinds of schematic patterns) in the formulation of concepts for speech. For example, I would suggest that using the schematic structures of image schemas in thinking-for-

speaking should involve less conscious effort and attention than using more complex, detailed spatial images. Therefore, the production of image-schematic forms gesturally would have to involve less conscious attention than the production of more elaborate hand shapes and movements. The kinds of gestures studied here – spontaneously produced co-verbal gestures – are what McNeill (1992: 72) has characterized as "unwitting accompaniments of speech." Therefore, there is a natural motivation for why the majority of the gestures studied in these conversations (A and O types) could, on the whole, be reliably characterized in terms of image schemas.[5] Furthermore, the fact that gestures display these kinds of schematic patterns that are shared in human experience can be a basis for their communicative potential. If language provides prompts for an addressee to construct mental simulations of the speaker's intended meaning (Barsalou 1999), some findings from research on non-human primates and humans suggest gesture's potential role in this process. Research on mirror neurons has shown that the same neurons are activated when a monkey observes an action and when the monkey performs the same action (Rizzolatti and Arbib 1998). Research with humans has shown that frontal motor cortex areas are activated not only when participants move their hands, but also when they observe other persons moving their own hands (DiPellgrino et al. 1992; Rizzolatti 1994). In other studies, it has been found that during hand action observation, but not in control conditions, there was an increase of amplitude of motor evoked potentials recorded from hand muscles which are normally used when actually performing an action (Fadiga et al. 1995). Further research has confirmed these findings and also found that the motor evoked potentials of the muscles were modulated in a fashion strictly resembling the time-course of the observed action (Gangitano et al. 2001). These and related findings lend support to the argument that observation of hand movements, such as gestures, can provide a richer context for constructing the simulations prompted by the co-occurring speech. The evidence in the present study shows one way in which gestures can provide other information than the speech alone — that is, by invoking additional or other schemas in terms of which to interpret the entity, action, or relation being talked about.

---

[5] On image schemas and mental imagery, see Gibbs and Berg 2002, Cienki 2002, and the other contributions in that issue of the *Journal of Mental Imagery*.

3.2 Future directions

Each condition tested for the present paper showed significant results in the use of image schemas as descriptors, but the latter analyses in the Results section revealed more agreement in the categorization of A gestures than O gestures, both without and with sound. The data suggest that the A gestures may have a more easily imageable schematic form than the O gestures. This could be because the A gestures are referential and so are more specific in form than the O gestures. In this regard it will be interesting to compare the results with those of the forthcoming study (Cienki, in prep.) on concrete referential gestures. It will help distinguish whether it is the referential nature of the gestures that allows them to be more readily characterized in terms of image schemas, or whether the abstract referential gestures constitute a special category. If the latter is the case, it will lend support to the argument that image schemas can play an important role as the source domains of many types of metaphors (Johnson 1987; Lakoff 1987), given that A gestures constitute a significant category of metaphoric gestures by virtue of depicting an abstract domain in terms of a concrete form (see Cienki and Müller, in prep.).

How do we reconcile the findings of these studies with what we know of the idiosyncratic nature of gesturing, namely that it is so variable across individuals, as noted by McNeill (1992)? While idiosyncratic in their execution, we can argue that (many or most) gestures have a basis in common patterns in our experience, and image schemas serve as prime examples of such patterns. Image schemas might provide common skeletal structures which underlie individuals' seemingly idiosyncratic gestures. For example, a given image schema can be manifested gesturally at different levels of complexity. Gestures thus not only reflect schematic patterns, abstracted from various domains of experience, they also instantiate them anew. Further research is needed on how commonly these patterns are shared cross-culturally; quantitative research in particular could shed light on whether some image schemas are interpreted as predominating more than others as the foundations of gesturing, within a given culture or across cultures.

Finally, let us consider the results of conditions 3 and 4, which were meant to be control conditions. How do we make sense out of why the

phrases without gestures were also reliably characterized with image schemas? As a number of papers in this volume make clear, the semantics of various kinds of words and constructions can be analyzed in terms of image schemas, so perhaps this finding should not be surprising. Other research points to a relation between the cognitive processing of gesture and speech. For example, it has been found that hand motor cortex is activated while simply listening to speech (Flöel et al. 2003). A question remains for the present experiment, however, as to whether there was a relation between the fact that the words/phrases could reliably be characterized with image schemas and the fact that the utterances were originally produced with gestures. This is particularly relevant for the phrases accompanying the A gestures, which, at least according to one measure reported in the Results, were categorized with greater agreement than the phrases accompanying O gestures. In future research it would be worth comparing whether phrases which were not accompanied by gestures could also be characterized reliably in terms of image schemas.

References

Barsalou, Lawrence W.
   1999   Perceptual symbol systems. *Behavioral and Brain Sciences* 22: 577-609.

Beattie, Geoffrey and Heather Shovelton
   1999   Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica* 123: 1-30.
   2001   An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture* 1: 129-149.

Bühler, Karl
   1982   *Sprachtheorie: Die Darstellungsfunktion der Sprache.* Stuttgart: Fischer.

Calbris, Geneviève
   2003   From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture* 3: 19-46.

Cienki, Alan
    1997   Some properties and groupings of image schemas. In: Marjolijn Verspoor, Kee Dong Lee, and Eve Sweetser (eds.), *Lexical and Syntactical Constructions and the Construction of Meaning*, 3-15. Amsterdam/Philadephia: John Benjamins.
    2002   Questions about mental imagery, gesture, and image schemas. *Journal of Mental Imagery* 26 (1&2): 43-46.
    2003   Ontological metaphors prevail in gesture with speech. Presented at the Eighth International Cognitive Linguistics Conference, La Rioja, Spain, July 2003.
    In prep.   Concrete versus abstract reference and the schematicity of gesture.

Cienki, Alan and Cornelia Müller (eds.)
    In prep.   *Metaphor and Gesture.* Amsterdam / Philadelphia: John Benjamins.

DiPellgrino, G., L. Fadiga, L. Foggassi, V. Gallese, and G. Rizzolatti
    1992   Understanding motor events. *Experimental Brain Research* 91: 176–180.

Eckman, Paul and Wallace V. Friesen
    1969   The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica* 1: 49-98.

Fadiga, L., L. Fogassi, G. Pavesi, and G. Rizzolatti
    1995   Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology* 73: 2608 –2611.

Flöel, A., T. Ellger, C. Breitenstein, and S. Knecht
    2003   Language perception activates the hand motor cortex: implications for motor theories of speech perception. *European Journal of Neuroscience* 18: 704-708.

Gangitano, M., F. M. Mottaghy, and A. Pascual-Leone
    2001   Phase-specific modulation of cortical motor output during movement observation. *Neuroreport* 12: 1489 –1492.

Gibbs, Raymond W., Jr. and Eric A. Berg
    2002   Mental imagery and embodied activity. *Journal of Mental Imagery* 26 (1&2): 1-30.

Johnson, Mark

1987   *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: University of Chicago Press.

Kendon, Adam
2004   *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Kita, Sotaro
2000   How representational gestures help speaking. In: David McNeill (ed.), *Language and Gesture*, 162-185. Cambridge: Cambridge University Press.

Kita, Sotaro (ed.)
2003   *Pointing: Where Language, Culture, and Cognition Meet.* Mahwah, NJ: Lawrence Erlbaum Associates.

Lakoff, George
1987   *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind.* Chicago: University of Chicago Press.

McNeill, David
1985   So you think gestures are nonverbal? *Psychological Review* 92: 350-371.
1992   *Hand and Mind: What Gestures Reveal about Thought.* Chicago: University of Chicago Press.

McNeill, David and Susan Duncan
2000   Growth points in thinking-for-speaking. In: David McNeill (ed.), *Language and Gesture*, 141-161. Cambridge: Cambridge University Press.

Müller, Cornelia
1998   *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich.* Berlin: Berlin Verlag Arno Spitz.

Rizzolatti, G.
1994   Nonconscious motor images. *Behavioral and Brain Sciences* 17: 220.

Rizzolatti, G. and M. Arbib
1998   Language within our grasp. *Trends in Neurosciences* 21(5): 188-194.

Slobin, Dan I.
   1987   Thinking for speaking. In: J. Aske *et al*. (eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 435-445. Berkeley, CA: Berkeley Linguistics Society.

Taub, Sarah F.
   2001   *Language from the Body: Iconicity and Metaphor in American Sign Language.* Cambridge: Cambridge University Press.

Appendix

Which type of stimuli in each condition were categorized more reliably using image schemas: A, O, or neither (—)?

Conditions	1 (gesture)	2 (gesture + sound)	3 (phrase + sound)	4 (phrase)
Mean # of image schemas chosen	A	A*	—	—
Less image schemas used at least once	—	A	—	—
All 7 descriptors used with fewer stimuli	A	—	A	—

*Marginally more reliably than O.