Cognitive skills and gesture–speech redundancy

Formulation difficulty or communicative strategy?

Autumn B. Hostetter and Martha W. Alibali
Kalamazoo College / University of Wisconsin-Madison

Speakers sometimes convey information in their gestures that they do not convey in the accompanying speech. The present study examined whether individual differences in the production of non-redundant gesture–speech combinations are related to individual differences in speakers' spatial and verbal skills. We classified speakers as spatial dominant, verbal dominant, or equally matched on the basis of the difference in their performance on a spatial visualization test and a verbal fluency test. We used the coding procedure developed by Alibali et al. (2009) to code speakers' gesture–speech redundancy as they narrated an animated cartoon. Spatial-dominant speakers produced a higher proportion of non-redundant gesture–speech combinations than other speakers. The results suggest that some speakers may use non-redundant gesture–speech combinations as a communicative strategy that enables them to capitalize on their strong imagistic representations.

Keywords: gesture, verbal skills, spatial skills, image, gesture–speech redundancy

Speakers communicate information in many ways. Most obviously, speakers often produce words and sentences that describe their entire meaning verbally. In addition, speakers produce nonverbal cues such as facial expressions, intonations, and hand gestures that can reinforce or even carry part of their meaning. For instance, imagine a speaker who says "he did some exercises" while pushing her hands outward from her body repeatedly, as though doing push ups. This gesture conveys more specific information about the kind of exercises she is describing than does her speech. Such gesture–speech combinations have been referred to as "mismatching" (e.g., Church & Goldin-Meadow, 1986), "complementary" (e.g., McNeill, 1992), or "non-redundant" (e.g., Alibali, Evans, Hostetter, Ryan, & Mainela-Arnold, 2009).

Gesture 11:1 (2011), 40–60. DOI 10.1075/gest.11.1.03hos
ISSN 1568–1475 / E-ISSN 1569–9773

The degree of redundancy between a gesture and its accompanying speech can be considered along a continuum (Goldin-Meadow, 2003). Because gestures and speech encode meanings in very different ways, all gestures are non-redundant with speech to some degree. The spatial, holistic medium of gesture supports the expression of more detailed spatial and motor information than does the verbal, linear medium of speech. Consider a speaker who describes the actions of a cartoon character on a high bar by saying "and he starts flipping around the bar" while quickly making three small circles with the index finger of her left hand. The gesture provides some information about how many flips occurred and how fast they were; however, the basic action event being described (e.g., spinning around the bar) is conveyed by both gesture and speech. In this sense, this gesture is largely redundant with the accompanying speech. In contrast, consider the "exercising" gesture described previously. This gesture conveys more information about the nature of the action being described than does the accompanying speech, and is therefore closer to the non-redundant end of the continuum. Throughout this paper, we use the term "non-redundant gesture" to refer to gestures that elaborate on the identity of the specific action or object that is referred to in a speaker's message. We use the term "redundant gesture" to refer to gestures that depict the same object or action that is mentioned in speech, even if they elaborate on some characteristic such as size, speed, or direction, as in the "flipping" example described above.

At first blush, dividing the information one wishes to convey between the verbal and gestural channels may seem like a less effective communicative strategy than conveying the entire message in speech. However, extensive evidence suggests that information that is conveyed uniquely in gesture is attended to and readily understood by listeners (e.g., Alibali, Flevares, & Goldin-Meadow, 1997; McNeill, Cassell, & McCullough, 1994). Indeed, a recent meta-analysis suggests that gestures that are non-redundant with speech actually have a greater influence on listener comprehension than gestures that are redundant with speech (Hostetter, 2011). Thus, producing non-redundant gestures with speech is a successful communicative strategy, and it may be as successful as articulating all of the important details in speech.

The purpose of the present study is to explore whether some people are more likely than others to produce non-redundant gesture–speech combinations, and more specifically, whether these individual differences are related to differences in individual speakers' cognitive skills. To understand why some speakers might produce more non-redundant gesture speech combinations than others, it is first necessary to understand why such combinations are produced at all. One likely possibility is that gesture and speech have different expressive possibilities. Because of their spatial, holistic nature, gestures are a natural choice for conveying spatial information. To quote Kendon (2004), gestures "can achieve adequate descriptions with much greater economy of effort and much more rapidly than words alone can manage" (p. 198). For example, a speaker who wishes to describe the precise angle at which a shape was tilted may have more communicative success showing the angle in gesture rather than trying to describe it in speech (e.g., Emmorey & Casey, 2001), because the spatial medium of gesture affords communication of this information much more readily than does the verbal medium of speech.

The expressive possibilities of gesture may be particularly valuable when the information being described is difficult for a speaker to lexicalize. Gestures occur more frequently with speech about figures that are difficult to name than with speech about figures that are easier to name (Graham & Heywood, 1975; Krauss & Hadar, 1999). In some situations, speakers use gestures to express important components of the meaning they wish to convey (Melinger & Levelt, 2004). Speakers may produce gestures when describing spatial information that is difficult to lexicalize because gestures are a natural way of representing the information that does not require that the difficult lexical items be retrieved.

It seems, then, that non-redundant gesture–speech combinations may be particularly likely when speakers have clear and specific spatial knowledge about what they wish to describe without a corresponding clear lexical label. According to this view, non-redundant gesture–speech combinations are not the simple result of strong spatial knowledge or of weak verbal knowledge, but they are instead the result of speakers having both strong spatial knowledge and weak verbal knowledge simultaneously. Indeed, the available evidence on when non-redundant gesture–speech combinations are produced is compatible with this hypothesis. We review this evidence below.

First, speakers produce non-redundant gestures when the linguistic possibilities (or lack thereof) of their language make it difficult for them to describe a particular idea (McNeill & Duncan, 2000). To illustrate, Japanese does not have an equivalent of the English verb *swing*. Kita and Özyürek (2003) compared the gestures produced by Japanese speakers to those of English speakers as they described a scene in which one character swings across a street on a rope. They found that 10 out of 15 Japanese speakers produced an arc shape in their gesture, although verbally they described "moving across", without explicitly noting the arc-shaped path. In this situation, it seems clear that the speakers have spatial knowledge of what they wish to convey (an arc-like trajectory) but they lack the ability to lexicalize it in their language. By producing arc-shaped gestures, Japanese speakers are able to express specific information about the nature of the action that is difficult for them to convey in their accompanying speech.

Second, children who are learning a new concept often convey information about the concept in their gestures that they do not also convey in their speech. For example, children learning about the concept of conservation sometimes focus on the height of one container in their speech, while focusing on the width of the container in their gesture (Church & Goldin-Meadow, 1986). Similarly, children who are learning about new scientific concepts often produce gestures about those concepts before they can accurately describe them in their speech (Roth, 2002; Singer, Radinsky, & Goldman, 2008). In these situations, children have some spatial knowledge of the content they are describing, but they may not have represented this information in verbal form, creating a situation in which their spatial understanding of the domain outstrips their verbal understanding. As a result, the information they express in gesture does not completely overlap with the information they express in words.

Third, speakers frequently produce non-redundant gesture–speech combinations when their verbal skills are still developing. Even when children completely understand the event they are describing and know exactly which aspects of their spatial knowledge they wish to articulate, their vocabularies may not contain the words they need to describe the precise spatial or motor properties of the events they are thinking about. Indeed, Alibali et al. (2009) found that children aged 5 to 10 produced non-redundant gesture–speech combinations while narrating a cartoon at a rate more than twice that of adults. Further, the children studied by Alibali et al. seemed particularly likely to produce non-redundant gestures when they were having trouble formulating their ideas into speech. As evidence, children produced more speech dysfluencies with non-redundant gesture–speech combinations than with redundant combinations. While this evidence may seem to point to a lack of verbal knowledge as the determining factor leading to non-redundant gesture–speech combinations, it may be the case that it is instead the combination of strong spatial knowledge and weak verbal knowledge. All of the spatial and motor events being described were relatively simple; thus, all children likely also had good spatial knowledge of the events.

The previous findings are compatible with the idea that non-redundant gesture–speech combinations occur when speakers have strong spatial knowledge combined with weak verbal knowledge; however, this hypothesis has not been put to test. The present study will test this hypothesis by directly measuring the discrepancy between speakers' spatial and verbal knowledge. Specifically, we propose that even adult speakers who are describing relatively simple spatial information that can be easily lexicalized in their language occasionally produce non-redundant gesture–speech combinations as a result of their characteristic pattern of thinking. Speakers have a choice about how to mentally encode and represent information, and there are at least two representational formats they may use: verbal and imagistic (e.g., Paivio & Sadoski, 2011). Which representational format a given speaker adopts for a given task depends on the demands of the particular task (Noordzij, van der Lubbe, & Postma, 2005), the instructions given (Mathews, Hunt, & MacLeod, 1980), and the speaker's individual preferences (Reichle, Carpenter, & Just, 2000). Importantly, speakers seem to choose a format that will minimize their own cognitive effort; participants who are good at visual-spatial tasks more frequently choose an imagistic format than a verbal format (MacLeod, Hunt, & Mathews, 1978). It seems likely, then, that when speakers with strong visual-spatial skills view an imagistic stimulus, such as a cartoon, they are likely to represent at least some of the events as visual images, rather than converting them to a verbal description. Furthermore, if a speaker with strong visual-spatial skills also has relatively weak verbal skills, he or she may be particularly unlikely to strongly activate a verbal description to correspond to particular aspects of the image. As a result, when the speaker describes the relevant mental representation, the details of the image may be strongly activated and easy to express in gesture, while the corresponding words may be less strongly activated and less accessible. Indeed, previous research suggests that speakers who have stronger visual-spatial skills than verbal skills produce higher rates of gestures than other speakers (Hostetter & Alibali, 2007). The purpose of the present study is to determine whether the gestures produced by such speakers are also particularly likely to convey non-redundant information.

Our central hypothesis is that speakers with stronger spatial skills than verbal skills will be more likely to produce non-redundant gesture–speech combinations than other speakers, because their rich mental images are more easily expressed in gesture than articulated in speech. These non-redundant gesture–speech combinations may reflect one of two communicative strategies. First, speakers may produce non-redundant gesture–speech combinations because they are attempting to convert their imagistic representations into a verbal code, but are having difficulty doing so. When speakers are having difficulties formulating or packaging speech, their speech typically becomes dysfluent, and these dysfluencies are frequently accompanied by gestures (Christenfeld, Schachter, & Bilous, 1991; Rauscher, Krauss, & Chen, 1996). Thus, it is possible that non-redundant gestures occur because the speakers are trying to package their imagistic representations into speech, but they do not readily find the precise lexical representations they are looking for. Under this view, speakers should be more dysfluent when producing non-redundant gesture–speech combinations than when producing redundant combinations.

In contrast, speakers may produce non-redundant gesture–speech combinations, not as a reflection of difficulties in accessing and formulating verbal codes, but because such combinations are a communicative strategy that allows them to capitalize on the different expressive possibilities of speech and gesture. By producing non-redundant gesture–speech combinations, speakers with clear mental images can avoid having to verbally encode all of the details in their speech. According to this view, non-redundant gesture–speech combinations occur because mental images are more active in speakers' minds at the moment of speaking than are verbal codes, regardless of whether speakers are actually attempting to access the verbal codes. According to this view, the speech produced with non-redundant gesture–speech combinations should not be particularly dysfluent.

In brief, the goal of the present investigation is to determine whether speakers whose spatial skill outstrips their verbal skill produce more non-redundant gesture–speech combinations than speakers with more equally matched or verbal-dominant cognitive profiles. We will also examine whether non-redundant gesture–speech combinations tend to involve dysfluent speech, in order to gain insight into the communicative strategies that give rise to gesture–speech non-redundancy.

To address these aims, we utilized the task and coding system developed by Alibali et al. (2009). This coding system assesses the redundancy of the gestures and speech produced by speakers retelling a cartoon about a mouse and an elephant. The system identifies the common meanings conveyed in gesture by speakers who are describing this particular cartoon, and specifies the commonalities in form of gestures that correspond to each particular meaning. Using the system, it is possible to code the meanings of gestures that are produced with descriptions of this cartoon *independently* from the accompanying speech. Such a system is important because if gesture and speech are coded together, the meaning of the speech may influence the interpretation of the gestures, making judgments of redundancy fairly subjective. For example, if a speaker produces the push-up gesture described at the outset of this paper along with the words "does some exercises," a coder might be likely to interpret the gesture as "exercise" (which would be redundant with speech) rather than "push up" (which conveys information not expressed in speech). In contrast, the system developed by Alibali et al. (2009) makes it possible to determine the meaning of each gesture without reference to the accompanying speech, and in turn, to make a more objective judgment of whether that meaning is redundant with the accompanying speech or whether it conveys additional information.

We applied this system for coding redundancy to the descriptions produced by speakers with three cognitive profiles (spatial dominant, verbal dominant, and equally matched), determined on the basis of their scores on spatial visualization and verbal fluency tests. We predicted that speakers with spatial-dominant cognitive profiles would produce more gestures that are non-redundant with the accompanying speech than would other speakers.

Method

Participants

The sample for this study ($N = 41$) was a sub-sample of the larger sample described in Hostetter and Alibali (2007).[1] All participants from the larger sample were recruited from the psychology research pool at the University of Wisconsin-Madison, and they received extra credit for participating. All were native speakers of English. Two considerations were made when selecting participants for inclusion in the subsample for the present study. First, we included only participants who produced at least one codable gesture while describing the cartoon stimulus (see below). Second, we chose participants on the basis of their cognitive profile. As described below, all participants completed measures of verbal fluency and spatial visualization. Each participant's score on each task was converted to a z-score based on the complete distribution of 90 scores from Hostetter and Alibali (2007). The difference between each participant's spatial visualization and verbal fluency z-scores was then used to classify his or her cognitive profile. Participants whose verbal skill was at least one standard deviation higher than their spatial skill were included as *verbal dominant* participants ($n = 14$; 8 female). Participants whose spatial skill was at least one standard deviation higher than their verbal skill were included as *spatial dominant* participants ($n = 12$; 6 female). A subset of participants whose verbal and spatial skills were within half a standard deviation of one another were randomly chosen as *equally matched* participants ($n = 15$; 10 female). In order to create cognitive profile groups that were as distinct as possible, we did not include participants whose skill sets were different by more than half but less than one standard deviation. The average age of individuals in the sub-sample analyzed here was 19.31 ($SD = 1.42$).

Procedure

The complete procedure for data collection is detailed in Hostetter and Alibali (2007). Briefly, participants arrived at the lab individually for a study about how people remember and communicate information. All participants completed a series of seven tasks in the same fixed order. Only the tasks relevant to the present analysis are discussed here.

First, participants completed a measure of verbal skill, known as the Controlled Oral Word Association Test (COWAT), Word Fluency, or the F-A-S test. The COWAT is a commonly used neuropsychological assessment tool that measures participants' skill at organizing and keeping track of lexical space (e.g., Martin, Wiggs, Lalonde, & Mack, 1994). For a full discussion of the COWAT and its relation to gesture, see Hostetter and Alibali (2007). In this test, participants had 60 seconds to name as many words as they could beginning with the letter 'S' and another 60 seconds to name as many words as they could beginning with the letter 'T'. The two trials were separated by the other tasks in the experiment and were not completed back to back. The total number of words named in each trial was summed, excluding morphological changes on the same word (e.g., "sit" and "sitting" did not count as different words) and proper nouns (e.g., Saturday and Sam). These totals were then averaged across the two trials to produce a verbal fluency score for each participant that was then converted to a z-score. Previous research (desRosiers & Kavanagh, 1987) has documented the retest reliability of the COWAT as .88.

Second, participants completed a measure of spatial visualization skill, known as the Paper Folding Test (Ekstrom, French, Harman, & Derman, 1976). This test measures participants' skill at mentally transforming an image. In this test, participants viewed a series of items in which a square piece of paper is folded along dotted lines and then punched with a hole. The participants then chose which of five choices correctly depicted what the paper would look like if it were unfolded. The test consisted of two pages with 10 items on each page. Participants had 3 minutes to complete each page, and there was no penalty for items they did not answer. Each participant's score was calculated as the total number correct (out of 20) minus one-quarter of the number incorrect. The test's creators estimated its retest reliability as approximately .76 (Ekstrom et al., 1976).

Third, participants watched a 90 s clip from the German cartoon series "Die Sendung mit der Maus." In the clip, a mouse, an elephant, and a leprechaun interact with a gymnastics high bar. A complete list of the events that occur in this cartoon is presented in Table 1. The cartoon is accompanied by music and sound effects but does not contain any spoken words. Participants viewed the cartoon twice on a Macintosh Powerbook G3 laptop with a 35 cm color screen. Following the second viewing, participants were asked to describe the events of the cartoon to the experimenter. Throughout the description, the experimenter maintained eye contact with the participant and reacted appropriately to the participant's narration (e.g., smiled during funny parts). The participants were told that their descriptions were being audio-recorded. A hidden video camera also recorded the participants' narratives and their accompanying gestures. Participants were informed about the presence of the video camera and the interest in gesture at the end of the experiment. All participants gave permission for their video data to be included in the analysis.

Table 1. Events in the cartoon, from Alibali et al. (2009)

1.	Mouse appears.
2.	Mouse walks up to bar.
3.	Mouse blinks.
4.	Mouse stands underneath bar.
5.	Mouse stretches / warms up.
6.	Mouse jumps up and grabs bar.
7.	Mouse swings.
8.	Mouse makes grunting noises.
9.	Mouse spins around bar.
10.	Elephant enters.
11.	Elephant watches mouse on bar.
12.	Mouse jumps off.
13.	Elephant "claps".
14.	Mouse turns around and sees elephant.
15.	Elephant walks up to bar.
16.	Elephant jumps, trying to get on bar.
17.	Elephant grabs bar with trunk.
18.	Elephant's trunk bends the bar.
19.	Mouse gets upset.
20.	Mouse scolds elephant.
21.	Elephant hangs for a couple seconds.
22.	Elephant gets off the bar.
23.	Mouse gestures for elephant to get back.
24.	Mouse tries to fix the bar.
25.	Leprechaun with top hat appears.
26.	Mouse gives up and steps back.
27.	Leprechaun walks under bar.
28.	Bar is magically repaired by leprechaun's hat.
29.	Leprechaun walks off.
30.	Elephant laughs.
31.	Mouse pouts and looks embarrassed.

Coding

One primary coder who was blind to the cognitive skill profile of each participant viewed the videotape of each cartoon description and transcribed the speech verbatim. All accompanying representational gestures were noted. *Representational gestures* are movements that depict semantic content. For example, a speaker who says "there was this gymnastics bar" while producing a bimanual motion in which the tips of the fingers touch in central space above her head and then move outward to both sides is producing a representational gesture. Individual representational gestures were segmented from one another based on changes in handshape or motion.

Following this first pass through the data, the primary coder then made a second pass through the data, watching each participant's description without access to the accompanying sound or transcript. On this pass, the coder identified representational gestures that were codable in the system developed by Alibali et al. (2009) for this cartoon. Alibali et al. identified 13 meanings that are commonly expressed in gesture by speakers narrating this particular cartoon, and they identified the commonalities in *form* shared by gestures that depict each meaning. The criteria for all 13 meanings can be seen in Table 2. For instance, Alibali et al. found that speakers who gesture while describing the event in which the mouse spins around the bar typically produce a repetitive circular motion with one or both hands. A gesture with this form can thus be coded as meaning SPIN, independently of the accompanying speech. Each representational gesture produced by speakers in this sample was considered for whether its form met the criteria for one of the meanings identified by Alibali et al. Approximately 54% of the representational gestures produced in this sample could be coded in this way.

Next, each codable gesture was considered for whether its meaning was redundant or non-redundant with the meaning of the accompanying speech, at both the clause and the word levels. To be redundant at the word level, the meaning conveyed by the gesture (e.g., SPIN) had to also be mentioned in speech at the exact moment that the gesture was produced (e.g., "the mouse was **spinning** around the bar" or "he did **some flips** on the bar"; the exact words coinciding with the gesture are indicated in bold). In contrast, a gesture was classified as non-redundant at the word level if the gesture's meaning was not expressed in the words immediately accompanying the gesture (e.g., "**and he tried** to go around the bar"). To be redundant at the clause level, the meaning conveyed in the gesture had to be conveyed in speech within the same clause in which the gesture was produced. All of the gestures in the preceding examples would fit this criterion. Gestures were classified as non-redundant at the clause level if the words in the accompanying clause never conveyed the meaning conveyed by the gesture (e.g., "and then he did some **gymnastics moves**"). Note that according to this system, all gestures that were redundant at the word level were also redundant at the clause level, and all gestures that were non-redundant at the clause level were also non-redundant at the word level. However, it was possible for gestures to be non-redundant with the exact words, but redundant with the clause, as in the "and he tried" example described above.

Finally, we also counted the speech dysfluencies that occurred with each gesture. We identified four types of speech dysfluencies: (1) *filled pauses* (e.g., um, uh), (2) *repetitions*, in which a word was repeated (e.g., around <u>the, the</u> bar), (3) *repairs*, in which the speaker changed a word within the same syntactic frame (e.g., the elephant <u>went, tried</u> to get on the bar), and (4) *restarts*, in which the speaker

Table 2. Gesture lexicon: Meanings and descriptions of gesture forms, from Alibali et al. (2009)

Gesture meaning	Description of gesture form
Swing	Gesture that includes a back and forth motion; may be produced with hands or legs
Spin	Gesture that includes a circular motion, typically repeated and in neutral space; may be produced with one or both hands
Bar	Gesture that traces or takes the form of the bar; one or both hands may point, flatten, or form O's to represent round shape of bar
Stand	Gesture that traces or takes the form of the stand; one or both hands (typically both) point or flatten to represent upright stand for bar
Bar + stand	Gesture that traces or takes the form of the stand; typically produced with both hands; points may trace shape of stand and bar, or hands (with fingertips together) bend at knuckles or wrists so fingers represent bar and palms or arms represent stand
Grab bar	Gesture in which hands hover in parallel, sometimes with grasping motion; typically produced with both hands, either in neutral space or above head
Bent bar	Gesture in which hands trace shape of bent bar *or* half of bent bar, *or* gesture in which hands hover while holding shape of bent bar; may be produced with one or both hands
Dismount	Gesture in which hand makes a downward arcing motion; may include a slight upward motion before the downward motion; typically produced with one hand
Hat	Gesture made on or above head, in which hands either trace hat shape, form hat shape with hands or point to (imaginary) hat; may be produced with one or both hands using either points or flat handshapes
Jump	Gesture in which hands move up and down several times; typically produced in neutral space or in lap; may be produced with one or both hands, using either flat or curved open handshapes
Push bar up	Gesture in which both hands, palms face up or out, move up; typically produced in high neutral space or above head
Up	Gesture in which one hand moves up, in either point, flat, or curved open handshape; typically produced in neutral space
Walk	Gesture that includes alternate stomping motion; can be produced with feet or hands

changed to a new syntactic frame in mid-utterance (e.g., <u>and then the, but</u> the mouse was angry).

Results

We hypothesized that speakers whose spatial skill is stronger than their verbal skill would be particularly likely to produce non-redundant gesture–speech combinations. To test this hypothesis, we entered all codable gestures into a mixed logistic regression with cognitive skill as a fixed factor and gesture meaning and participant as random factors. The dependent variable was whether each gesture was redundant or not with the accompanying speech. In the first analysis, we considered redundancy at the word level. In the second analysis, we considered redundancy at the clause level.

In the analysis of redundancy at the word level, the model that included cognitive profile explained significantly more variance than a model that included only participant and gesture meaning as random factors, $\chi 2(2, N=227)=6.28$, $p=.04$. Specifically, there was a significant effect of the spatial-dominant cognitive profile (see Figure 1). The odds that a gesture produced by a speaker with a spatial-dominant cognitive profile was non-redundant with the accompanying words were 2.32 times higher than the odds that a gesture produced by an equally-matched speaker was non-redundant with the accompanying words, $\beta=0.844$, $SE=.41$, $z=2.04$, $p=.04$. Similarly, the odds that a gesture produced by a speaker with a spatial-dominant cognitive profile was non-redundant with the accompanying speech were 2.29 times higher than the odds that a gesture produced by a verbal-dominant speaker was non-redundant with the accompanying words, $\beta=0.827$, $SE=.35$, $z=2.36$, $p=.02$. There was no difference in the odds that a gesture would be non-redundant with the accompanying words between speakers with an equally-matched cognitive profile and those with a verbal-dominant cognitive profile, $\beta=0.02$, $SE=.44$, $z=0.04$, $p=.97$.

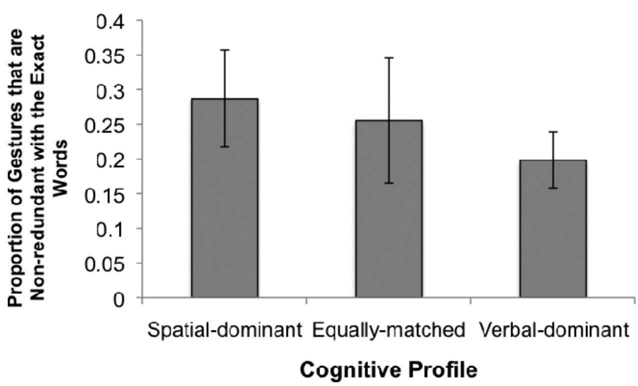

Figure 1. The average proportion of representational gestures that were non-redundant with the exact words in the accompanying speech, across the three cognitive profiles. Error bars represent standard errors of the means.

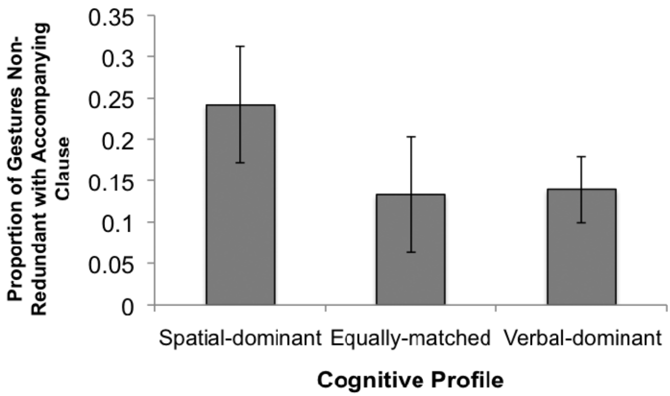

Figure 2. The average proportion of representational gestures that were non-redundant with the accompanying clause, across the three cognitive profiles. Error bars represent standard errors of the means.

The analysis of redundancy at the clause level mirrored the analysis at the word level (see Figure 2). The model that included cognitive profile explained more variance than a model that included only the random factors of participant and gesture meaning, $\chi 2(2, N=227)=5.78$, $p=.055$. As in the word level analysis, there was a significant effect of the spatial-dominant cognitive profile. The odds that a gesture produced by a spatial-dominant speaker was non-redundant with the accompanying clause were 3.63 times higher than the odds that a gesture produced by an equally-matched speaker was non-redundant with the accompanying clause, $\beta=1.29$, $SE=.58$, $z=2.23$, $p=.03$. The odds that a gesture produced by a spatially-dominant speaker was non-redundant with the accompanying clause were 2.56 times higher than the odds that a gesture produced by a verbal-dominant speaker was non-redundant with the accompanying clause, $\beta=0.94$, $SE=.47$, $z=2.01$, $p=.04$. As in the word level analysis, there was no difference in the odds that a gesture was non-redundant with the accompanying words between speakers with an equally-matched cognitive profile and those with a verbal-dominant cognitive profile, $\beta=0.35$, $SE=.60$, $z=0.58$, $p=.56$.

Next we considered the frequency of speech dysfluencies. We entered the number of dysfluencies speakers produced during each gesture–speech combination into a mixed Poisson regression model with the number of words produced in the clause as an offset variable. Poisson regression predicts the frequency of some event (in this case, frequency of speech dysfluenices). By including an offset variable, we are modeling the rate of the event relative to that offset variable (in this case, speech dysfluencies per word). The redundancy of the combination (redundant with exact words or not), the cognitive profile of the speaker (spatial, equal, or verbal), and the interaction between redundancy and cognitive profile

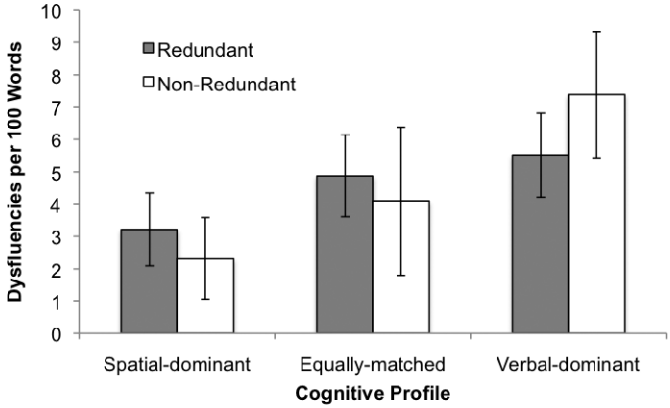

Figure 3. The average dysfluency rates per 100 words produced with gestures that were redundant and non-redundant with the accompanying words, as a function of the speaker's cognitive profile. Error bars represent standard errors the means.

were included as fixed factors. Gesture meaning and participant were included as random factors. This model explained more variance in the rate of dysfluencies produced than a model that included only gesture meaning and participant as random factors and word count as an offset variable, $\chi 2(5, N=227)=18.96, p=.002$.

The results are presented in Figure 3. Verbal-dominant speakers produced more dysfluencies per word with non-redundant gesture–speech combinations than with redundant gesture–speech combinations, $\beta=3.65$, $SE=.86$, $z=4.26$, $p<.001$. In contrast, spatial-dominant speakers and equally-matched speakers produced comparable rates of dysfluencies per word with redundant gesture–speech combinations and with non-redundant combinations, $\beta=-0.17$, $SE=.78$, $z=-0.22$, $p=.83$ for spatial-dominant speakers and $\beta=-0.07$, $SE=.78$, $z=-0.09$, $p=.93$ for equally-matched speakers. The pattern shown by verbal-dominant speakers differed significantly from the patterns shown by equally-matched speakers, $\beta=3.73$, $SE=1.16$, $z=3.22$, $p=.001$, and spatial-dominant speakers, $\beta=3.82$, $SE=1.21$, $z=3.16$, $p=.002$. There was no difference in the patterns shown by equally-matched and spatial-dominant speakers, $\beta=0.10$, $SE=1.09$, $z=0.09$, $p=.93$.

Finally, we considered each gesture–speech combination that was non-redundant at the clause level for whether the information conveyed in gesture was described in speech at some other point in the narrative. In 64% of the cases in which a gesture was non-redundant with the speech in the accompanying clause, the unique information depicted in gesture was never described in speech. To determine whether the likelihood of eventually describing the unique information in speech was dependent on cognitive profile, we analyzed non-redundant clauses in a mixed logistic regression with cognitive profile as a fixed factor and participant and gesture meaning as random factors; the dependent variable was whether or not the gestured meaning was expressed in speech elsewhere in the narrative. There was no effect of cognitive profile on the likelihood of describing gestured information in speech elsewhere in the narrative; the model that included cognitive profile did not explain more variance than a model that included only participant and gesture meaning as random factors, $\chi2(2, N=47)=2.86$, $p=.24$.

Discussion

We predicted that there would be individual differences in how frequently speakers convey information in gesture that they do not also convey in the accompanying speech. Specifically, we hypothesized that these individual differences would be related to individual differences in speakers' proficiency with different types of mental representations. This hypothesis was supported by the present data. Speakers with a spatial-dominant cognitive profile produced a higher proportion of gesture–speech combinations that were non-redundant than did speakers with verbal-dominant or equally matched cognitive profiles.

Further, there was no evidence to suggest that the increased production of non-redundant gesture–speech combinations by spatial-dominant speakers was due to increased difficulties with speech production. Spatial-dominant speakers were not more dysfluent with non-redundant than redundant clauses, suggesting that they were not having particular trouble with speech production when they produced non-redundant gesture–speech combinations. Instead, it appears that they produce non-redundant gesture–speech combinations as a specific communicative strategy that allows them to express their imagistic representations without translating them into a verbal code.

This conclusion contrasts with the conclusion drawn by Alibali et al. (2009) regarding children's increased production of non-redundant gesture–speech combinations relative to adults. Unlike the spatial-dominant adult speakers in the present analysis, the children studied by Alibali et al. produced more dysfluencies overall than adults, and they were particularly likely to produce dysfluencies with non-redundant gesture–speech combinations. Taken together, the present results and those of Alibali et al. (2009) suggest that speakers who have a clear image of what they want to describe without strongly activated corresponding verbal codes are likely to produce gestures that uniquely convey some of their imagistic knowledge. In some cases, speakers may produce these non-redundant gestures as they attempt to access and formulate an appropriate verbal message, as the children in Alibali et al. appeared to do. In other cases, such as the spatial-dominant adults described here, speakers may produce non-redundant gestures as part of a broader communicative strategy that enables them to capitalize on their imagistic representations without having to access corresponding verbal codes.

Unlike the spatial-dominant speakers, the verbal-dominant speakers in this study tended to produce speech dysfluencies more frequently when they produced non-redundant gesture–speech combinations than when they produced redundant combinations. Thus, it seems that the verbal-dominant speakers may have used a different communicative strategy than the spatial-dominant speakers. Specifically, it seems that the verbal-dominant speakers sought to express *in speech* all of the information that they wish to communicate. These speakers are usually successful at lexicalizing all of the spatial information they wish to express, so most of their gestures are redundant with speech. However, when they are unsuccessful, this is manifested both in speech dysfluencies and in non-redundant gesture–speech combinations. Spatial-dominant speakers, in contrast, may make less effort to express all of the relevant information in speech. As a consequence, they speak more fluently and also more often express aspects of their spatial mental representations in gestures that are not redundant with speech.

Regardless of the cognitive profile of the speaker, in the majority of instances, the information conveyed by gesture in non-redundant gesture–speech combinations was never described explicitly in speech. For example, at the end of the cartoon, the leprechaun's hat bumps the bent bar from underneath and magically pushes it up so that it is straight again. Many speakers describe this event by saying something like "his hat fixes it" while producing a gesture that clearly conveys the meaning UP (e.g., one hand, palm facing up, bends upward from the wrist). This gesture is providing more specific information than the speech about *how* the bar was fixed. However, very few speakers went on to describe this upward motion explicitly in their speech.

These findings contrast with those of Alibali et al. (2009), in which participants in the adult sample resolved approximately 75% of non-redundancies in a subsequent speech clause, compared to only 34% in the present sample. The source of this discrepancy is unclear, but it may be due to the fact that we included many participants with unmatched cognitive profiles in the present study. We examined the cognitive profiles of the adults who produced gestures in the Alibali et el. (2009) dataset, and found that the majority (66%) had equally-matched cognitive profiles, compared to only 37% of the present sample. It is possible that the greater representation of speakers with uneven cognitive profiles in the present study may be the reason for the lower percentage of resolutions, compared to the sample in Alibali et al (2009).In the present study, the likelihood of resolving non-redundant gestures in speech did not depend on cognitive profile; however, the pattern of means is suggestive, and such an effect might be revealed in a study with greater statistical power. In any case, further investigation of this issue is warranted.

Why do speakers sometimes avoid expressing information in speech that has already been conveyed in gesture? One possibility is that from the speaker's perspective, the information conveyed in gesture has already been communicated, making it unnecessary to articulate it in speech as well. This view aligns with the claim that speakers intend for their gestures to communicate (e.g., Kendon, 2004; Melinger & Levelt, 2004), and suggests that speakers treat the information conveyed in gesture as part of the common ground they have created with their listeners (Gerwing & Bavelas, 2004).

A second possibility is that speakers do not consider the information they convey uniquely in gesture to be particularly important, and that this is why they often choose not to articulate the information explicitly in speech. Of course, such judgements hinge on the specifics of the communicative task that speakers are faced with. In the example described above, perhaps speakers do not see the details of how the leprechaun fixed the bar to be important in this particular communicative situation; indeed, it is only the fact that he did so that leads to the next event in the story (the elephant laughing). More research is needed to determine whether the information conveyed uniquely in gesture is considered by speakers to be as important as the information they convey in speech.

The present findings suggest that speakers who have stronger spatial skills than verbal skills are particularly likely to use their gestures to convey information that they do not convey in their speech. We claim that this happens because spatial-dominant speakers are likely to have strongly activated mental images of the information they are describing and less strongly activated corresponding verbal codes. While tenable, this claim is based on two assumptions that should be tested more directly in future research. First, the present measures of cognitive skills are only indirect measures of how likely a speaker is to mentally represent events imagistically, and the present analysis rests on the assumption that speakers typically choose to represent information in the format with which they are the most proficient. Although this assumption is supported by previous research (e.g., Reichle et al., 2000), the use of a more direct measure in future research would be preferable. Second, speakers with all three cognitive profiles produced some non-redundant gesture–speech combinations. It is our contention that this is because all speakers sometimes represent information imagistically while speaking; however, future research is needed to determine what factors other than an individual's cognitive profile might predict whether speakers represent information imagistically or verbally at a particular moment while speaking.

Two other limitations to the present study should also be noted. Participants in the sample studied here were selected because they conformed to the particular cognitive profiles we were interested in. The extent to which the present findings extend to participants with less extreme differences in cognitive skills remains to be determined. Finally, it must also be noted that the results here pertain to a specific set of gestures that could be coded independently of the accompanying speech. Indeed, the analysis addresses only the subset of participants' gestures (about 50%) that were codable within the system developed by Alibali et al. (2009). Furthermore, some participants were excluded because they did not produce any gestures at all while describing the cartoon.[2] Thus, the extent to which these findings generalize to gestures with other meanings produced in other circumstances is unclear.

Despite these limitations, the present results are clear: speakers whose spatial skills outstrip their verbal skill produce more non-redundant gesture–speech combinations than other speakers. This finding can be explained by at least two contemporary theories about how gestures arise. First, it is compatible with Growth Point theory (McNeill, 1992, 2005). According to this view, gestures arise from the imagistic components of "growth points," or minimal units of thought. A growth point consists of both imagistic components and linguistic categorical components; as the growth point is described, the imagistic components come to be expressed as speech-accompanying gestures during speaking. The present data suggest that the contrast between imagistic and linguistic components of thought may be stronger for some speakers (i.e., spatial-dominant speakers) than others, leading them to more frequently produce gestures that depict the image of their growth point without simultaneously expressing the linguistic component in speech.

Second, the findings are also compatible with the Gesture as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). The GSA framework proposes that gestures occur as outward manifestations of the cortical motor activity that is involved in thinking about imagistic events. Simulations, or re-enactments of neural activity that were involved in the encoding of an event, activate motor and premotor cortex during speech about the event. When this motor activation is sufficiently strong, it is realized as an overt gesture alongside speech. According to this view, gestures are reflections of the mental simulations that occur with thoughts about imagistic events. Images evoke stronger motor activation than do verbal representations, and are therefore more likely to result in gesture. Thus, it is possible that speakers who tend to represent events imagistically as they are speaking are particularly likely to gesture about those events. Further, such speakers may not activate verbal codes for some components of those imagistic events, leading speakers to produce non-redundant gesture–speech combinations.

In sum, the present study provides the first evidence that there are clear individual differences in the semantic overlap of gesture and speech, and that these individual differences are related to speakers' profiles of verbal and spatial skills. These findings add to the growing knowledge base about the factors that lead to individual differences in gesture production. Moreover, the present findings suggest that non-redundant gesture–speech combinations may not be due to formulation difficulty, but instead may reflect a communicative strategy that is useful for conveying rich spatial information. As such, the findings underscore the fact that gestures are an integral part of successful communication — perhaps more so for some speakers than others.

Notes

1. Twelve of the participants included here were also included in the adult sample ($N=20$) analyzed in Alibali et al. (2009).

2. Such participants could not be included here because it is not possible to calculate a proportion of gestures that are redundant with speech when they produced zero gestures. However, these participants are included in the analysis reported in Hostetter and Alibali (2007) which focuses on gesture rate rather than on gesture redundancy.

References

Alibali, Martha W., Julia L. Evans, Autumn B. Hostetter, Kristen Ryan, & Elena Mainela-Arnold (2009). Gesture–speech integration in narrative: Are children less redundant than adults? *Gesture*, 9, 290–311.

Alibali, Martha W., Lucia Flevares, & Susan Goldin-Meadow (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology, 89*(183–193).

Christenfeld, Nicholas, Stanley Schachter, & Frances Bilous (1991). Filled pauses and gestures: It's not coincidence. *Journal of Psycholinguistic Research*, 20, 1–10.

Church, Ruth Breckinridge & Susan Goldin-Meadow (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition*, 23, 43–71.

DesRosiers, Gabriel & David Kavanagh (1987). Cognitive assessment in closed head injury: Stability, validity, and parallel forms for two neuropsychological measures of recovery. *International Journal of Clinical Neuropsychology*, 9, 162–173.

Ekstrom, Ruth B., J. W. French, Harry H. Harman, & Diran Derman (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Emmorey, Karen & Shannon Casey (2001). Gesture, thought, and spatial language. *Gesture*, 1, 35–50.

Gerwing, Jennifer & Janet B. Bavelas (2004). Linguistic influences on gesture's form. *Gesture*, 4, 157–195.

Goldin-Meadow, Susan (2003). *Hearing gesture: how our hands help us think*. Cambridge, MA: Harvard University Press.

Graham, Jean Ann & Simon Heywood (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, 5, 189–195.

Hostetter, Autumn B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137, 297–315.

Hostetter, Autumn B. & Martha W. Alibali (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, 7, 73–95.
Hostetter, Autumn B. & Martha W. Alibali (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15 (3), 495–514.
Kendon, Adam. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.
Kita, Sotaro & Aslı Özyürek (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of thinking and speaking. *Journal of Memory and Language*, 48, 16–32.
Krauss, Robert. M. & Uri Hadar (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, speech, and sign* (pp. 93–116). Oxford, UK: Oxford University Press.
MacLeod, Colin M., Earl B. Hunt, & Nancy N. Mathews (1978). Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior*, 17, 493–507.
Martin, Alex, Cheri L. Wiggs, François Lalonde, & Carol Mack (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks. *Neuropsyhchologia*, 32, 1487–1494.
Mathews, Nancy N., Earl B. Hunt, & Colin M. MacLeod (1980). Strategy choice and strategy training in sentence-picture verification. *Journal of Verbal Learning and Verbal Behavior*, 19, 531–548.
McNeill, David (1992). *Hand and mind: what gestures reveal about thought.* Chicago: The University of Chicago Press.
McNeill, David (2005). *Gesture and thought*. Chicago: The University of Chicago Press.
McNeill, David, Justine Cassell, & Karl-Erik McCullough (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, 27, 223–237.
McNeill, David & Susan Duncan (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge, UK: Cambridge University Press.
Melinger, Alissa & Willem J. M. Levelt (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4, 119–141.
Noordzij, Matthijs L., Rob H. J. van der Lubbe, & Albert Postma (2005). Strategic and automatic components in the processing of linguistic spatial relations. *Acta Psychologica*, 119, 1–20.
Paivio, Alan & Mark Sadoski (2011). Lexicons, contexts, events, and images: Commentary on Elman (2009) from the perspective of Dual Coding Theory. *Cognitive Science*, 35, 198–209.
Rauscher, Frances, Robert M. Krauss, & Yihsiu Chen (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7, 226–231.
Reichle, Erik D., Patricia A. Carpenter, & Marcel Adam Just (2000). The neural bases of strategy and skill in sentence-picture verification. *Cognitive Psychology*, 40, 261–295.
Roth, Wolff-Michael (2002). From action to discourse: The bridging function of gestures. *Cognitive Systems Research*, 3, 535–554.
Singer, Melissa, Joshua Radinsky, & Susan R. Goldman (2008). The role of gesture in meaning construction. *Discourse Processes*, 45, 365–386.

Authors' addresses

Autumn B. Hostetter
Kalamazoo College
Department of Psychology
1200 Academy Street
Kalamazoo, MI 49006
USA

autumn.hostetter@kzoo.edu

Martha W. Alibali
Department of Psychology
University of Wisconsin-Madison
1202 W. Johnson Street
Madison, WI, 53705
USA

mwalibali@wisc.edu

About the authors

Autumn B. Hostetter is Assistant Professor in the Psychology Department at Kalamazoo College. Her research examines the cognitive underpinnings of speech-accompanying gestures, particularly as they relate to theories about the embodiment of cognition. She is also interested in the sources of individual differences in both producing and comprehending gestures.

Martha W. Alibali is Professor of Psychology at the University of Wisconsin, Madison. Her research investigates mechanisms of knowledge change in cognitive development and learning, with a special focus on the role of gesture in processes of change. She is also interested in the role of gesture in communication, particularly in educational settings. She is the author of more than 40 journal articles and book chapters, and her work has been or is currently funded by the National Science Foundation and the Institute of Education Sciences.