Relations Between Observer Viewpoint Gestures and Visual Processing Abilities in Preschoolers

Abstract

According to the gesture-as-simulated-action (GSA) framework, simulated visual imagery produces observer viewpoint gestures, whereas simulated motor imagery produces character viewpoint gestures. Although this relation has been reported for adults, little is known about whether it holds for children. Therefore, we conducted a study with 4-year-old children (M = 50 months, SD = 3.4) and hypothesized that children with higher visual processing (*Gv*) abilities and children with higher fluid intelligence (*Gf*) would engage more in visual imagery simulation and, therefore, perform a higher rate of observer viewpoint gestures than children with lower *Gv* or *Gf* abilities. In the first session, we observed gestures in 39 children across three different communicative tasks. In the second session, we administered a SON-R test to assess children's cognitive abilities. Results revealed strong associations between the frequency with which children used the observer viewpoint in their co-speech gestures and their *Gv* abilities, but no relation between observer viewpoint gestures and *Gf*. Because we found this result in all three communicative tasks, we assume it is a general rather than a task-specific phenomenon. We discuss our results in terms of how *Gv* abilities, and thus observer viewpoint gestures, help children to achieve communicative goals.

Keywords: GSA framework, iconic gestures, viewpoints, cognitive abilities, mental images

Introduction

Research on gestures reveals a large and growing body of literature on the underlying cognitive mechanisms (e.g., Abramov et al., 2021; Chu & Kita, 2008; Hostetter & Alibali, 2008), and there are many theories arguing that the origin of gestures is to be found in mental processes (e.g. Kita & Özyürek, 2003). In this article, we focus on the gesture-as-simulated-action (GSA) framework. This proposes that gestures are tightly coupled to motor and perceptual processes that are likely to be reflected in a speaker's simulated mental imageries (Hostetter & Alibali, 2008, 2019). In this embodied view, when persons reason about objects and events that are currently not present perceptually, they activate similar sensorimotor processes to those that would be involved in actually performing the action or viewing the scene (e.g., Alibali & Kita, 2010; Chu & Kita, 2011; Rizzolatti, Fogassi, & Gallese, 2001; Wilson, 2002). How strongly simulated mental imageries are activated is related to individual differences (see for review: Özer & Göksun, 2020). In their review, Özer and Göksun (2020) point out that most research focuses on children of different ages and on how different verbal skills relate to children's gesture production (Austin & Sweller, 2014; Church, Kelly, & Lynch, 2000; Perry, Church, & Goldin-Meadow, 1992). However, which cognitive abilities drive spontaneous gesture production in early childhood is still unknown (Özer & Göksun, 2020). Our study addresses this research gap by investigating how individual differences in cognitive abilities relate to children's iconic co-speech gestures. Iconic co-speech gestures bear semantic information about objects and actions and have been reported to be generated from mental images. We focus specifically on observer viewpoint gestures. These are iconic gestures depicting trajectories of objects from a third-person perspective (McNeill, 1992; Parrill, 2010). The GSA framework assumes that observer viewpoint gestures originate from visual mental imagery (Hostetter & Alibali, 2008). We tested this assumption by relating children's visual processing (*Gv*) abilities and fluid intelligence (*Gf*) to their use of observer viewpoint gestures. *Gv* abilities refer to the ability to manipulate, recall, and think with visual stimuli, whereas Gf refers to the ability to solve novel problems, particularly those of a spatial nature (Cattell, 1963; Schretlen et al., 2000). Importantly, both *Gv* abilities and *Gf* have been reported to be involved in the process of visual mental imagery generation from which observer viewpoint gestures are assumed to emerge (Flanagan & Dixon, 2014; Hostetter & Alibali, 2008, 2019; Lohman, 1988; Sassenberg, Foth, Wartenburger, & van der Meer, 2011).

Gestures and Mental Imagery

Various studies have investigated how mental images influence people's gesture behavior. Hostetter and Skirving (2011), for example, provided participants with an additional visual or oral presentation of a story and found that in the process of speaking, speakers' co-speech gestures are associated with the perceptual stimuli with which they have experienced an event. Participants who watched and then listened (without pictures) to a story produced more gestures during a retelling task than those who listened to the story twice. This suggests that visual stimuli generate richer mental images, resulting in a greater frequency of gestures during retelling (Hostetter & Skirving, 2011). Other studies have investigated whether a stronger activation of mental imagery during speaking results in a greater production of gestures. Accordingly, Smithson and Nicoladis (2014) asked people to retell a cartoon story while viewing an unrelated visuospatial array that was either complex or simple. The authors found that speakers gestured more while watching the more complex visuospatial array, suggesting that demanding visual input led to an activation of mental imagery that boosted gesture production (Smithson & Nicoladis, 2014). Similarly, Sassenberg and van der Meer (2010) let participants explore a differently complex visuospatial array and studied their descriptions of easy and complicated routes. They found that participants produced more gestures while watching a complex visuospatial array during an easy route's description than when describing a complicated route while experiencing a simple visuospatial array.

These findings emphasize that the activation of mental images is an essential part of gesture generation. However, whereas these results indicate that the activation's strength can be influenced externally, it is plausible to argue that it can also be influenced internally by a person's individual cognitive abilities. This argument resonates with findings on individual differences and their relation to the frequency of gestures that speakers perform. Hostetter and Alibali (2007) found that speakers with low verbal but high visualization abilities used representational gestures more frequently than participants with other constellations of visual and verbal skills. This result suggests a possible connection between verbal and spatial skills: When people with low verbal skills but high spatial visualization skills communicate, they seem to draw on their more pronounced spatial abilities, and this ongoing processing increases their gesture frequency (Hostetter & Alibali, 2007). Along these lines, Ehrlich et al. (2006) asked 5-year-old children to explain how they solved a spatial transformation task. They found that it was the children's rate of gestures expressing movements that correlated positively with their spatial abilities. Whereas previous studies have reported on the relation between spatial abilities and gesture production in order to explain individual differences in gesture frequency (Ehrlich et al. 2006; Hostetter & Alibali, 2007), the study by Sassenberg et al. (2011) associated gestures with the more general fluid intelligence (*Gf*)—that is, the ability to understand complex relationships and to solve novel reasoning problems (Cattell, 1963).

In their study, Sassenberg et al. (2011) found that compared to students with average *Gf*, students with high *Gf* solved complex spatial tasks faster and more accurately while utilizing more gestures that express movements from an observer perspective. As mentioned above, observer viewpoint gestures are iconic gestures that depict trajectories of objects. The authors concluded that young adults with high *Gf* engage more in visual imagery than their peers with average *Gf*. Hence, the findings of this study indicate a relation between gesture production and a more general cognitive ability—fluid intelligence (Sassenberg et al., 2011).

Overall, the literature reviewed above supports the multicomponential processes involved in generating mental images that seem to rely on situational constraints on the one hand and the individual's cognitive abilities on the other. However, the multiple components in this process also concern the kind of mental images that are generated within the production of specific iconic gestures. In the next section, we shall follow up on this component.

Kinds of Mental Images: Viewpoints in Gestures

Recall that the GSA framework proposes that two kinds of gestures need to be differentiated (Hostetter & Alibali, 2008, 2019): Whereas the character viewpoint is suggested to be related to motor imagery—as if the speaker performed the action her- or himself—the observer viewpoint—that is, the way the speaker is watching the events as an outside observer—is related to stimulated visual imagery (Hostetter & Alibali, 2008 p. 504; McNeill, 1992). Nonetheless, Parrill (2010) points out convincingly that the viewpoint someone takes when gesturing relates strongly to the corresponding aspects of an event. For example, think of a ball that is thrown and moves through the air until a character catches it with her or his hands. There are aspects of this event that require different descriptions, and these, in turn, elicit different viewpoints in gestures. In her study, Parrill (2010) found that when speakers used gesture to depict a character who is using her or his hands, it was mostly character viewpoint gestures that were elicited, whereas the reference to a movement's trajectories elicited observer viewpoint gestures. More concretely, Parrill (2010) observed that speakers exclusively used character viewpoints in their gestures when referring to someone catching a ball (a motion that requires an agent to use the hands). Referring to this aspect of the event, one speaker mimicked the hands of an observed character and pretended to catch a ball. Mimicking a character's hand is defined as *hand-as-hand* gestures (Cartmill, Rissman, Novack, & Goldin-Meadow, 2017). In contrast, aspects referring to a flying ball elicited exclusively observer viewpoint gestures (Parrill, 2010). Performing such a gesture, a speaker would model the referent (the ball) with her or his hand—which is described as *hand-as-object* gestures (Cartmill et al., 2017). However, it has to be noted critically that defining viewpoints by whether the gesturing hand is used as object or hand is problematic. Cartmill et al. (2017) made an important observation that in addition to hand-as-hand gestures, there are specific cases in which character viewpoint gestures can also be performed with hand-as-object gestures. This is the case when the speaker's body is oriented toward the gesture in such a way that suggests a character viewpoint, but the handshape is hand-as-object (Cartmill et al., 2017, p. 44)—for example, depicting how to eat with a spoon while modeling the spoon's shape with the hand (see Table 1). We shall return to this issue when discussing our results.

A further study by Parrill and Stec (2018) examined how persons talk about pictures and provided evidence suggesting that exogenous factors such as event structure are closely related to a gesture's viewpoint. The authors analyzed the viewpoint used in gestures according to whether they were performed from a first- or third-person view, and they found that speakers who experienced events from the first-person perspective used character viewpoint gestures at a higher rate than participants who experienced events from a third-person perspective.

A study by Sassenberg et al. (2011) also supplements the findings on endogenous factors (cognitive abilities) contributing to mental images: Investigating young adults with high *Gf*, the authors found that when explaining how to solve an analogy task, they performed more gestures from an observer perspective than their peers with average *Gf*. Nonetheless, further supportive evidence for other settings and groups of participants is lacking.

Whereas the studies reported so far provide evidence for a relation between mental images and the viewpoint in gesture, another function of viewpoints in gestures that has been discussed recently relates to discourse structure (Debreslioska & Gullberg, 2019; Demir, Levine, & Goldin-Meadow, 2015; McNeill D., 1992; Parrill, 2010). For example, Debreslioska and Gullberg (2019) suggest that the viewpoint in gesture functions as a cohesive device in narrative discourse. More specifically, in a narration, it is necessary to identify “who did what to whom” (Stites & Özçalışkan, 2017, p. 1029). Within this identification of crucial story elements during storytelling, people tend to use character viewpoint gestures if a referent is maintained, whereas they typically perform observer viewpoint gestures to reintroduce a referent. Hence, the gesture’s viewpoint is considered to be an indicator of the accessibility of a character (Debreslioska & Gullberg, 2019). The intriguing point here is that a perspective expressed in the use of gestures might be linked to a specific communicative task: the narration. In this vein, and when investigating developmental effects of the use of viewpoints, Demir et al. (2015) have analyzed how far the use of a particular viewpoint early in the development of narrative competencies predicts later narrative performance. They found that reenacting observed characters by performing character viewpoint gestures at the age of 4 aids children in structuring a retold story more accurately at a later age (Demir et al., 2015). The authors explained that character viewpoint gestures are likely to reflect empathy with observed characters. This empathy might lead to first-person knowledge and, therefore, to a more detailed recall of events (Demir et al., 2015). Beyond this proposed explanation, it is interesting to note that individual differences in the use of a particular viewpoint were predictive of later narration skills. This effect clearly speaks to individual differences requiring further investigation in terms of how the use of viewpoints might be related to what kind of cognitive skills.

The Present Study

The following study aims to investigate how individual differences in cognitive abilities relate to the viewpoint in iconic gestures from a developmental perspective. To address this question, we first related children's visual processing (*Gv*) abilities to their use observer viewpoint gestures. *Gv* abilities are responsible for perceiving, analyzing, synthesizing, manipulating, recalling, and thinking with visual patterns and stimuli. This includes spatial relationships, visual memory, and length estimation (Flanagan & Dixon, 2014; Lohmann, 1994). Importantly, *Gv* abilities are involved in forming visual mental images (Flanagan & Dixon, 2014) from which, according to the GSA framework, observer viewpoint gestures arise (Hostetter & Alibali, 2008). With this in mind, we assume that the more pronounced a child's *Gv* abilities, the more enriched visual mental images that child will generate, and this will lead to a higher rate of observer viewpoint gestures. Second, we aimed to explore the relation between gestures and the more general cognitive ability of fluid intelligence (*Gf*) reported by Sassenberg et al. (2011) further and investigate it in children performing other tasks. As mentioned above, Sassenberg et al. (2011) showed that when talking about solving a geometric analogy task, adults with high *Gf* performed more gestures that express movements from an observer perspective than adults with average *Gf*. From this, we hypothesized that children with higher *Gf* would perform observer viewpoint gestures at a higher rate than children with lower *Gf*. Hence, we can use our study to test another group of participants (children at the age of 4 years) when performing different communicative tasks. Relating children's cognitive abilities to their use of gestures across different tasks allows us to argue that there are specific or general underlying cognitive mechanisms that lead to the performance of observer viewpoint gestures. The tasks in our study differ in their content, which is crucial, because iconic gestures contain the semantic content of the referent to which they refer (McNeill, 1992). Please note, the tasks were designed for a project which focus on children's iconic gestures in different communicative genres. Each genre imposes its own demands, and children need to organize their talk appropriately (Labov & Waletsky, 1973; Mandler, 1984, Quasthoff et al., 2017). In this article, we use task synonymously for genre.

To sum up, research in recent years has focused on how children's individual differences influence their gesture production. Despite the growing interest in this research area, it is still widely unknown how children's gesture behavior is associated with their cognitive abilities. Our study aimed to relate children's observer viewpoint gestures to their *Gv* and *Gf* cognitive abilities, because these have both been reported to be involved in visual mental image generation. The relation of the investigated cognitive abilities to visual mental images is important because, according to the GSA framework, observer viewpoint gestures arise from visual mental imagery. However, it is also crucial to note that not only such endogenous factors as cognitive abilities but also exogenous factors are involved in visual mental image generation and, thus, in gesture production. With this in mind, we elicited children's gestures in three tasks that differed in their exogenous factors. Analyzing the relation of children's cognitive abilities to their gesture behavior in various tasks allows us to argue that there are general or more specific underlying cognitive mechanisms.

Method

Participants

Fifty-five preschool children from the regions of Bielefeld and Paderborn (North-Rhine Westphalia, Germany) participated in this study. Their ages ranged from 45 to 61 months (M = 50 months, SD = 3.4). Based on a pilot study, we considered 4 years to be the earliest age at which children's iconic gestures as co-speech gestures can be elicited reliably and under comparable conditions. Sixteen children had to be excluded for the following reasons: a deviation from the procedure specified for the communicative task of retelling (7 children), and failure to schedule the second appointment at which the intelligence test SON-

R was administered (9 children). Children received a book or DVD after the first appointment as reimbursement for their participation and a small toy after the second appointment.

Procedure

The study consisted of two sessions. Each was scheduled on a separate day within one month but at least two weeks apart. Both appointments were in the lab, and the children were invited together with their caregiver.

In accordance with Bielefeld University's ethics procedures for research with children, parents provided written consent to their children's participation at the first session. The children also provided verbal consent before participating. Additionally, they were informed that they could break off the interaction at any time.

At the first session, children performed on three communicative tasks (explanation, retelling, and illustration) in which both the experimenter and the caregiver were engaged. Each task was designed under two conditions to investigate the influence of possible exogenous factors. In the following, however, we shall focus on children's gesture behavior without considering the different conditions, because we found no significant differences in the children's proportions of viewpoints in gesture between the conditions. Children were assigned to one of the conditions before the start of the first session (see below for a more detailed description of the communicative tasks).

At the second appointment, we administered the nonverbal intelligence test SON-R (Tellegen et al., 2007). This test aims to measure children's visual processing abilities (*Gv*) and fluid intelligence (*Gf*) (Mickley & Renner, 2010, 2019).

Communicative Tasks

During the communicative tasks, children interacted with a caregiver, because an extensive pilot study in which we explored the conditions under which children gesture the most revealed that interaction with a caregiver allowed children to communicate in a familiar way. This elicited a more natural gesture behavior that we expected would reflect the children's ability more accurately. Indeed, at this age, when communicating, children are often scaffolded in their verbal productions. Caregivers do this in a fine-tuned way. In all communicative tasks, we followed a method of providing input first and eliciting communication afterward. More specifically, children first experienced an event with the experimenter while caregivers waited outside the room where they were given some instructions to read. Afterwards, the children could then communicate the event to their caregivers. The experimenter's role was thus to provide the child with experiences worth telling or to initiate the conversation between child and caregiver. The experimenter used a script for this to achieve better comparability across participants. We should highlight that we controlled the caregiver's behavior only very slightly by providing some written explanations and prompts. These suggestions were applied immediately before the interaction commenced. In total, three communicative tasks were designed to assess children's gestural behavior: explanation, retelling, and illustration.

Explanation. The study followed a fixed order of communicative tasks starting with explanation. After a warm-up time and filling out the consent form, the experimenter asked the parent to leave the room and told the child that they would be playing a game. Then, the experimenter introduced a self-made jigsaw puzzle to the child and explained the rules. Two conditions were designed for the experimenter's explanation: In one condition, the experimenter explained the game mostly verbally; in the other condition, children received the same verbal introduction supported by iconic gestures (see SUPPLEMENTARY ITEMS). The original aim of these conditions was to investigate possible alignment effects (Bergmann

& Kopp, 2012). As for the goal of this game, the child was told that all the pieces of the puzzle needed to be removed from the board. The puzzle pieces depicted parts of a little town at night on the board (see SUPPLEMENTARY ITEMS). The puzzle pieces could be removed by a Playmobil© figure "flying" over them. To make the figure fly, the child was allowed to throw a dice six times. The sides of the dice represented the different shapes of the little town and the sky. After the child threw the dice for the sixth time, the experimenter initiated the end of the game with the question, "Has the figure flown everywhere?" If the child answered "yes," the experimenter replied, "That's great!" If the child answered "no," the experimenter replied "No? Well, maybe next time!" After finishing the game, the experimenter put it away and asked the caregiver to reenter the room. They then all sat down on cushions on the floor together. The experimenter asked the child to explain the game to the caregiver so that she or he could play it. After successfully initiating the children's explanation, the experimenter left the room.

Retelling. The retelling task was prepared one day before the family arrived at the lab. Families were contacted via post one to two days before they came to the lab and sent a German version of a commercially available book or DVD with the story of a mole ("The mole and the green star" by Doskočilová et al., 1998/2013). In the letter accompanying either the DVD or the book, families were instructed to watch the movie or read the book one day before visiting us in our lab. Importantly, we asked that another caregiver than the one who would be visiting the lab should participate in the home activity with the child. This gave a valid pragmatic justification for the child to tell the story to the caregiver who visited the lab.

To initiate the retelling task, the experimenter reentered the room after hearing that the child had completed the explanation (the first task). The child and caregiver remained in the same sitting constellation. Then, the experimenter initiated the retelling by saying "The game that we just played reminds me of the star found by a mole. Do you know this story? Did you see/read the story with your [another caregiver]? Could you retell it for your mother/father?"

After the child started her or his retelling from the book or DVD, the experimenter left the room again. The motivation for having two different conditions (book/video) was to investigate whether visual stimuli evoke more gestures than verbal stimuli (Hostetter & Hopkins, 2002).

Illustration with examples. Finally, after hearing that the child had finished her or his retelling, the experimenter reentered the room and asked the caregiver to wait outside. Then, the experimenter and the child sat down at a table. There, the experimenter used a hand puppet (a dog) to perform some actions. A cover story was used for this performance with the experimenter introducing the dog as a knowledgeable animal that is proud to demonstrate what it knows about humans because it has already been living with them for a long time. In total, the dog demonstrated five daily situations. However, in each case, the dog did something inappropriate. This inappropriateness was surprising and, therefore, funny to the children. For example, the dog demonstrated how to eat from a plate with a spoon but held the spoon the wrong way round. After each demonstration, the dog asked, “Do I know/did I do it right?” Usually, the children were amused and were eager to correct the dog. Children’s corrections were designed to follow two conditions: In one condition, children were encouraged to stand up, walk to the dog, and show how to perform the actions correctly. In the other condition, children were asked to remain seated and to describe the appropriate actions verbally. The two conditions aimed to explore whether a performed action affects how children gesture about this event afterward. In this case, the condition in which children were allowed to walk over to the dog and to enact the action directly should stimulate their later gesture production. After the five events, the experimenter put the dog away, and the caregiver was invited to reenter the room. While waiting outside, the caregiver had been given some written instructions on what type of question to ask the child (see SUPPLEMENTARY ITEMS). Again, the caregiver and child sat down on the cushions on the floor, and the experimenter encouraged the child to report what she or he just experienced with the dog by asking, for example, “Can you tell your father/mother what the dog did?” Then, the experimenter left the room.

Stimuli

Explanation. For the explanation, a game board, a Playmobil figure, and a dice with five forms and a blank side on it were used. The game was designed to contain puzzle pieces with the same shapes as the dice (star, rectangle, triangle, circle, moon) that would give rise to many depicting gestures. In one condition, the game was explained to the child with iconic gestures; in the other, without iconic gestures. In the gesture condition, the experimenter performed a total of 7 iconic gestures: For all shapes on the board game (5) as well as for flying through a shape on the board game and for not allowing the child to fly with the figure (see, for a more detailed description, Kern, (2020)).

Retelling. For retelling, one condition used a movie; the other, a book. Both media are commercially available and about the same story: “The mole and the green star.” The film lasts about 7 min with colored moving pictures but no speech. The book contains selected static pictures of the movie, but most of the story is narrated in continuous text. The story is about a mole who finds a green star. He tries to pin the star back to the sky with the help of other animals. After several attempts have failed, the star is stolen by a magpie. Once the mole has retrieved the star, he finally manages to attach it to the sky with the help of the moon.

Illustration with examples. All of the five actions that the experimenter performed with the hand puppet are summarized in the SUPPLEMENTARY ITEMS. For these actions, the experimenter acted on additional objects and followed a fixed order reported in the SUPPLEMENTARY ITEMS.

Intelligence Test

At a second session, approximately two weeks after the first, we administered the nonverbal intelligence test SON-R (Tellegen, Laros, & Petermann, 2007). For the children's comfort, their caregiver was continuously present in the room but outside the child's field of view. To ensure comparability, caregivers were instructed not to interfere in the test situation.

The SON-R intelligence test is designed to measure visuospatial abilities (subscale PS IQ) and abstract and concrete reasoning (subscale RS IQ). According to the Cattell–Horn–Carroll (CHC) theory, the subscale PS IQ assesses visual processing, whereas the subscale RS IQ assesses fluid intelligence (*Gf*) (Mickley & Renner, 2010, 2019). During the testing procedure, feedback from the experimenter is accepted to a certain extent, because the interaction between tester and child is considered to be behavior providing a natural environment for children to demonstrate their cognitive abilities (Laros & Tellegen, 2015).

Table 1

Structure of the non-verbal intelligence test SON-R (Mickley & Renner, 2010, 2019; Tellegen et al., 2007)

Subscales	Test items	Measures
Visual Processing (PS IQ)	Drawing patterns	Visuo-motor skills; action planning; spatial thinking; (the ability to perceive and reproduce spatial position and the arrangement of a figure in a differentiated way)
	Mosaics	Thinking in spatial relationships (ability to grasp form relations between parts and the whole and act according to the template and synthesis of the individual parts)
	Puzzles	Analytical and synthetic thinking (perception of space-situation and figure-ground relationships, but also the child's environmental experiences)
Fluid Reasoning (RS IQ)	Situations	Concrete reasoning in concrete situations
	Categories	Abstract thinking and derive principles of order, grouping objects into categories according to common characteristics
	Analogies	Abstract thinking and deductive thinking (recognizing and applying sorting principles, recognizing and reproducing analogies and regularities)

Coding

Speech. Within the explanation, retelling, and illustration tasks, children's speech was transcribed and segmented into intonation phrases (Selting et al., 2009). Intonation phrases were used to control for the children's verbosity (number of gestures divided by intonation phrases), because they were not given a time limit resulting in a wide variation in the number of utterances produced. To account for this variation, we divided the number of gestures that the children used by the number of intonation phrases they produced.

Gesture. We identified iconic gestures using the taxonomy given by Cartmill et al. (2017) who categorize iconic gestures into *hand-as-hand*, *hand-as-object*, and *hand-as-neutral* gestures:

1. Hand-as-hand gestures occur when an action is mirrored (e.g., the action of brushing teeth is performed as if somebody were holding and using a toothbrush).

2. Hand-as-object gestures stand for an object (e.g., the index finger is standing for the toothbrush and the performer moves it as if she or he were brushing the teeth).

3. Hand-as-neutral (also called tracer gestures) are performed with a pointing finger but provide some symbolic information by drawing the shape or the movement in the air.

We applied this taxonomy for three reasons: First, it was developed to assess behavior in children (in contrast to adults). Second, the taxonomy helps to identify semantic information in hand movements and, therefore, it distinguishes iconic gestures from other types and hand movements. Third, this taxonomy aided us in our second coding step, which was to identify the gesture's viewpoint.

Within iconic gestures, four different viewpoints can be differentiated (McNeill, 1992): the character viewpoint (C-VPT), the observer viewpoint (O-VPT), the dual viewpoint, or no viewpoint. The taxonomy of "*hand-as-*" gestures is similar to that of viewpoints in gesture, but not identical. For example, hand-as-hand gestures are generally considered character viewpoint gestures, because both gesture types reflect how to handle an object. However, *hand-as-object* gestures can also be used from the first-person perspective—for example, if the speaker models a spoon with her or his hands and pretends to eat with this spoon. In contrast to first-person perspective gestures, observer viewpoint gestures are mostly *hand-as-object* gestures and *hand-as-neutral* gestures (Parrill, 2010). Gestures are considered observer perspective gestures when they depict the trajectory of an object's movement. A typical example of an observer viewpoint gesture is to show how an object falls. *Hand-as-neutral* gestures are considered observer viewpoint gestures when they depict an object's trajectory without modeling the object (Parrill, 2010).

Table 2

Possible Hand Types With Which a Specific Viewpoint in Gesture Can Be Performed

	Viewpoints	
	Character viewpoint	Observer viewpoint
Hand type	Hand-as-hand	Hand-as-object
	Hand-as-object	Neutral (drawing)

The 'no-viewpoint' type of gestures are performed with *hand-as-neutral* gestures and mostly depict the shape of an object. The dual-viewpoint type of gestures represents gestures expressing the O-VPT and the C-VPT simultaneously. However, because no-viewpoint and dual-viewpoint gestures rarely occurred in our data, we excluded these categories from further analysis.

Two independent coders assessed the reliability of coding different viewpoints in gesture on 10% of the data. Using Cohen's Kappa (Cohen, 1960) to measure interrater reliability, we found a substantial agreement for viewpoints within gestures of k = .79 (Landis & Koch, 1977). More specifically, there was an agreement of k = .83 for character viewpoint gestures and k = .75 for observer viewpoint gestures.

Data Analysis

We analyzed children's gesture behavior in two steps. First, we applied two separate repeated measures ANOVAs. One ANOVA was conducted with the independent variable "communicative tasks" (explanation, retelling, illustration) in order to test effects on iconic gesture frequency. For the second ANOVA, we used the variable "communicative tasks" as the independent variable, while testing for effects on the rate of character and observer viewpoint gestures. Greenhouse–Geisser corrections were applied where necessary. Significant interaction effects were resolved by Bonferroni-corrected post hoc pairwise comparisons.

Second, we conducted several linear regressions with children's achieved scores for *Gv* and *Gf* as independent variables and children's frequency of iconic, character viewpoint and observer viewpoint gestures as dependent variables.

Results

In the following, we shall first report on children's gesture behavior within and between the three communicative tasks—explanation, retelling, and illustration—before moving to the analysis of children's cognitive abilities.

Communicative Tasks

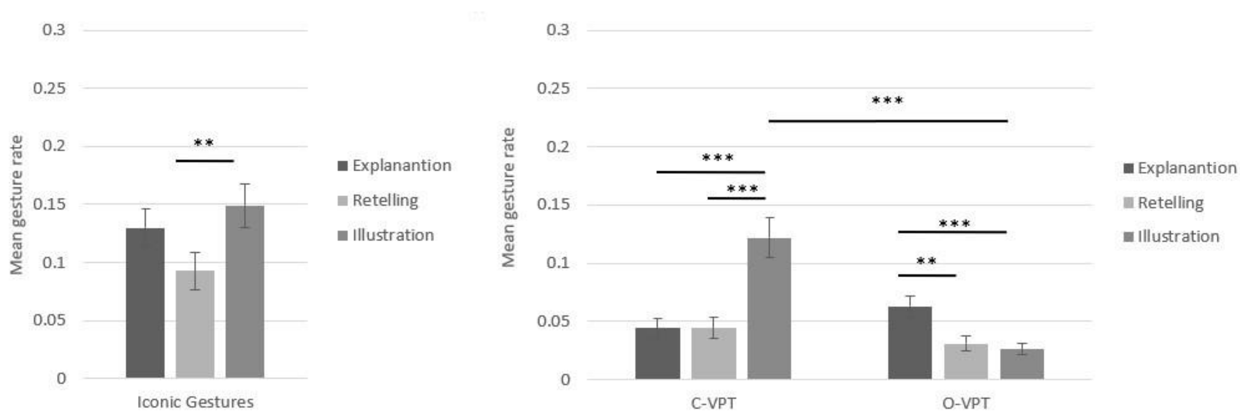
We conducted a repeated measures ANOVA with the proportion of iconic gestures as the dependent variable and the three different communication tasks (explanation, retelling, illustration) as independent variables to analyze children's gestures in the communicative tasks. There was a significant intermediate effect for the task variable, $F(2, 76) = 4.74$ $p < .05$, $\eta^2 = .11$, indicating that the proportion of iconic gestures differed according to tasks. Bonferroni-corrected post hoc pairwise comparisons revealed that iconic gestures were performed significantly more frequently in the illustration task than in the retelling task ($p < .01$; see Figure 1).

Next, we analyzed whether the viewpoint children adopted in iconic gestures differed between and within tasks. A repeated measures ANOVA with two viewpoints (character viewpoint vs. observer viewpoint) as dependent variables and the three tasks (explanation, retelling, and illustration) as independent variable revealed a significant interaction effect, $F(2, 76) = 20.13$, $p < .01$, $\eta^2 = .35$, indicating that there was a different rate of specific viewpoints in gesture between and within tasks. Bonferroni-corrected post hoc pairwise comparisons revealed that character viewpoint gestures occurred more frequently in the illustration task than in explanation and retelling ($p < .01$). With respect to observer viewpoint gestures, Bonferroni-corrected post hoc pairwise comparisons revealed a higher rate of observer viewpoints in explanation than in retelling and illustration ($p < .01$). Turning to differences within tasks, Bonferroni-corrected post hoc pairwise comparisons showed that character viewpoint gestures were more frequent than observer viewpoints in the illustration task ($p < .01$). No differences in the proportion of a specific viewpoint in gesture were found in either explanation or retelling (see Figure 1).

Figure 1

Mean Gesture Rates (SE) for (A) Iconic Gestures and for (B) Character Viewpoint (C-VPT) and Observer Viewpoint (O-VPT) Gestures Within the Explanation, Retelling, and Illustration Tasks

(A) (B)



Note. Number of gestures divided by intonation phrases. * $p < .05$. ** $p < .01$. *** $p < .001$.

In summary, we found considerable variance in the proportion of gestural viewpoints distributed both across and within tasks. The explanation task elicited proportionally more observer viewpoint gestures than the retelling and illustration tasks. Character viewpoint gestures occurred proportionally more often in the illustration task than in other tasks. Moreover, children performed proportionally more character viewpoint gestures than observer viewpoint gestures in the illustration task.

Gestures and SON-R Scores

In the next step, we analyzed whether children's cognitive abilities related to their gesture behavior by relating children's scores on the subscales of the nonverbal intelligence test SON-R to their proportions of viewpoints adopted in iconic gestures. The SON-R consists of the subscales SON-PS (visual processing) and SON-RS (fluid intelligence). We analyzed whether the subscales predicted the variance in character viewpoint and observer viewpoint gestures.

Visual Processing

To investigate relations between children's *Gv* (visual processing) score and their gesture production, we conducted several linear regression analyses using children's achieved scores within the subscale for visual processing (SON-PS) as independent variable and the proportion of iconic gestures as well as character viewpoint and observer viewpoint gestures as dependent variables (see Table 3). First, we related children's *Gv* scores to their rate of iconic gestures. We found no effect within all three tasks, indicating that children's Gv score could not predict the variance in children's proportions of iconic gestures. Zooming into the gestural viewpoints, we applied a linear regression and used observer viewpoint gestures as a dependent variable (see Table 3). This revealed that children's visual processing scores significantly predicted their rate of proportionally used observer viewpoint gestures within all three tasks. More specifically, within the explanation task, 14.9% of the variation in children's use of observer viewpoint gestures could be explained by children's *Gv* score, $F(1, 37) = 6.46$, $p < .05$. Within the retelling task, 29% of the variation in observer viewpoint gestures could be explained by children's *Gv* score, $F(1, 37) = 15.08$, $p < .001$. Finally, within the illustration task, 10% of the variation in children's use of observer viewpoint gestures was explained by children's *Gv* score, $F(1, 37) = 4.15$, $p < .05$. In contrast, the linear regression analyses for character viewpoint did not yield statistically significant predictions.

In sum, we found no general relation between children's *Gv* abilities and the occurrence of iconic gestures. However, separating iconic gestures into different viewpoints yielded the result that children's *Gv* abilities predicted the proportional use of observer viewpoint in all three tasks.

Fluid Intelligence

To investigate how children's fluid intelligence (*Gf*) related to their iconic gesture behavior, we conducted linear regressions using children's scores on the subscale for *Gf* as an independent variable. The dependent variable was, first, children's rate of iconic gestures in general and, second, the two viewpoints in gestures. Results showed that children's fluid intelligence did not predict their iconic gesture behavior (see Table 3).

Table 3

Relation of Children's Visual Processing (Gv) and Fluid Intelligence (Gf) to Iconic Gestures in General and Specifically to Character and Observer Viewpoint Gestures

Explanation

	Iconic		C-VPT		O-VPT	
Predictor	*R*	R^2 - Change	*R*	R^2 - Change	*R*	R^2 - Change
Gv	.29	.09	.14	.02	.38	.15*
Gf	.08	.01	.10	.01	.14	.02

Retelling

	Iconic		C-VPT		O-VPT	
Predictor	*R*	R^2 - Change	*R*	R^2 - Change	*R*	R^2 - Change
Gv	.30	.09	.16	.03	.54	.29***
Gf	.11	.01	.09	.01	.16	.03

Illustration

	Iconic		C-VPT		O-VPT	
Predictor	*R*	R^2 - Change	*R*	R^2 - Change	*R*	R^2 - Change
Gv	.22	.05	.13	.02	.32	.10*
Gf	.19	.04	.10	.01	.31	.09

$p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

The gesture-as-simulated-action (GSA) framework proposes that the viewpoints adopted in iconic co-speech gestures relate to simulated mental images. Specifically, GSA proposes that observer viewpoint gestures arise from underlying visual imagery, whereas character viewpoint gestures arise from motor imagery (Hostetter & Alibali, 2008). In this article, we investigated how cognitive abilities as endogenous factors influence gestures in terms of the viewpoint children apply in their iconic co-speech gestures. We did this by conducting a study in which we related children's visual processing abilities (*Gv*) and fluid intelligence (*Gf*) to their use of observer viewpoint gestures. We carefully designed three communicative tasks: explanation, retelling, and illustration. In these tasks, children interacted with their caregivers to elicit a natural communication behavior. To assess children's *Gv* and *Gf* abilities, we administered a nonverbal intelligence test during a separate session.

First, we examined the relationship between children's visual processing (*Gv*) abilities and children's use of observer viewpoint gestures. To remind the reader, *Gv* abilities refer to the ability to generate, store, and manipulate visual mental images (Flanagan & Dixon, 2014). We hypothesized that children with higher *Gv* abilities would engage more with simulated visual mental images leading to a higher rate of observer viewpoint gestures during communication. Confirming our hypotheses, linear regressions showed that children's Gv abilities were associated positively with the frequency of observer viewpoint gestures within all three communicative tasks. This means that the higher children's *Gv* abilities, the higher the expected frequency of their observer viewpoint gestures. We have to highlight the fact that— despite crucially different tasks with respect to children's speech production, the requirements to involve mental imagery, and the scaffolding that is possible by caregivers— the relation persisted across the different communicative tasks, yielding a quite general mechanism that is applied in all of them. To the best of our knowledge, our study is among the first to investigate possible differences in different forms of discourse systematically as a basis with which to account for specific or more general cognitive task requirements.

Second, we aimed to extend Sassenberg et al.'s (2011) findings suggesting that young adults with high fluid intelligence (*Gf*) use more gestures from an observer perspective than speakers with average fluid intelligence. We hypothesized that children with higher *Gf* would engage more strongly in visual imagery than their peers with lower *Gf*, leading to a higher rate of observer viewpoint gestures. Our results do not confirm this assumption. As a result, we suggest that the relation between *Gf* and observer viewpoint is a task-specific phenomenon or applies only to adults.

Our analyses reveal that the children's rate of observer viewpoint gestures increased linearly with children's *Gv* abilities in all three tasks. This finding allows us to reason that *Gv* abilities relate to the activations of simulated visual mental images from which observer viewpoints arise (Hostetter & Alibali, 2008). The tasks' situational constraints and events differed strongly in our study, which is reflected in how the tasks varied in eliciting specific viewpoints in gesture (see Figure 1). The fact that children's *Gv* abilities could be associated with observer viewpoint gestures within all three tasks indicates that this relation is a general rather than a task-specific phenomenon. While speaking, children engage with respect to their *Gv* abilities in visual mental simulations that they express in observer viewpoint gestures. Additionally, it indicates that a child's current gesture threshold can be surpassed more easily with a higher activation level (Hostetter & Alibali, 2008). According to the GSA framework, the production of a gesture depends not only on the strength of a mental simulation but also on the current gesture threshold. The gesture threshold is conceptualized as the speaker's current resistance to producing a gesture and is considered to be variable. The actual level of the threshold depends on dispositional and situational factors (Hostetter & Alibali, 2008). Our results within all three tasks indicate that the higher the activations of mental simulations the more likely the current threshold will be surpassed, leading to gesture production.

Overall, our results align with research arguing that the form and functions of gestures originate not only from exogenous factors but also from thinking processes as an endogenous factor (e.g., Chu & Kita, 2011; Ehrlich, Levine, & Goldin-Meadow, 2006). Depending on their *Gv* abilities, children seem to simulate visual mental imagery with sufficient strength to surpass their gestures threshold and thus produce observer viewpoint gestures (Hostetter & Alibali, 2008; 2019). Although we found strong evidence for this relation, we did not investigate how it affects children's communication. One possibility is that children with higher *Gv* are more likely to depict aspects of an event in gesture from an observer perspective than from a first-person perspective. For example, to show how someone climbs a ladder, the speaker could depict the upwards movement of the character in the gesture (observer viewpoint) instead of showing the actual movements of the hands for climbing (character viewpoint). In other words, an event that can be performed in gesture from a character or an observer viewpoint is more likely to be performed from an observer viewpoint when a child's *Gv* is high. Another reason that children with higher *Gv* perform more observer viewpoint gestures might be that they choose to talk more about particular aspects of an event, and these aspects predominantly elicit observer viewpoint gestures. Aspects of events that elicit observer viewpoint gestures refer to trajectories of objects with no motor actions involved (Parrill, 2010). It seems reasonable that children with higher *Gv* recall such spatial events more efficiently than events in which actions of the body are in focus. Consequently, children with higher *Gv* address more spatial events involving trajectories of objects than events with motoric content in their speech, and this, in turn, leads to a higher frequency of observer viewpoint gestures.

Another function of observer viewpoint gestures is suggested by Cartmill et al. (2012). This function is linked to children's cognitive load, because visuospatial mental simulations do not have to be retained in working memory when gesturing (Cartmill, Beilock, & Goldin-Meadow, 2012). Thanks to gestures, visuospatial mental simulations can be projected to an external space providing external visual cues that can be used to keep task-related visuospatial information active in working memory. This mechanism is considered to decrease the speaker's cognitive load (Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Along these lines, Cartmill et al. (2012) argue that especially observer viewpoint gestures are likely to decrease a speaker's cognitive load because the object is represented by the speaker's hand (hand-as-object gesture). This form of representation allows for some cognitive offloading, because the object does not need to be represented mentally. In contrast, when performing character viewpoint gestures, objects are mostly presented imaginatively with a hand-as-hand gesture. This form of representation implies that the imaginary object will be maintained in the speaker's mind (Cartmill et al., 2012). Following this argumentation, it seems reasonable that children who produce observer viewpoint gestures at a higher rate due to their pronounced *Gv* abilities free up more cognitive resources than children with lower *Gv* abilities. Thus, performing observer viewpoint gestures might benefit children in reaching communicative goals in two ways: On the one hand, children provide the listener with task-relevant information through observer viewpoint gestures; on the other hand, they can devote more cognitive effort into recalling and structuring the content of the given task. It should be noted, however, that due to our coding schema, children also performed character viewpoint gestures in which the object is modeled by the speaker's hand (e.g., modeling a spoon and pretending to eat with it). This point is critical in terms of not only how different viewpoints in gesture might decrease a child's cognitive load but also from what kind of mental images a gesture emerges. The GSA framework assumes that character viewpoint gestures arise from simulated motoric mental images. This implies a view in which character viewpoint gestures are always performed with (and defined by) hand-as-hand gestures, because hand-as-hand gestures reflect a hand's motoric behavior. In our view, however, the speaker's conceptual perspective on aspects of an event cannot be determined by analyzing the speaker's hand type alone (hand-as-hand vs.

hand-as-object). It is also critical to include the speaker's gaze, body orientation, and how the hands move in relation to the speaker's body (Cartmill et al., 2017; Frederiksen, 2017; Stec, 2012). In this vein, Cartmill et al. (2017, p. 44) argue that character viewpoint gestures are gestures in which "the gesture is located in space using the body as a frame of reference in a way that suggests a character viewpoint, but the handshape is hand-as-object."(Cartmill et al., 2017, p. 44). Investigations that blend gestures ("character viewpoint gestures" that are performed with hand-as-object gestures) and character viewpoint gestures with hand-as-hand gestures are functionally similar and belong to the same cognitive mechanism are currently lacking in the relevant literature. Further studies that take the different hand types into account will need to investigate how cognitive abilities and, therefore, mental images relate to character viewpoint gestures.

Our second hypothesis was motivated by Sassenberg et al.'s (2011) study. They found that students with high *Gf* used more gestures from an observer perspective than students with average *Gf* while explaining strategies to solve analogical reasoning tasks. Individuals with high fluid intelligence perform well on such tasks (Raven, 1958; Vernon, 1983). The proposed explanation states that people with high *Gf* are assumed to focus very efficiently on task-relevant information (Sassenberg et al., 2011). Along these lines, Sassenberg et al. (2011) argued that young adults with high *Gf* focused more on the rotational movements of the object during the task. This was indicated in their hand gestures during their explanation. However, our second hypothesis suggesting that children with higher fluid intelligence (*Gf*) would perform a higher rate of observer viewpoint gestures than children with lower *Gf* could not be confirmed. One explanation for our result is that in contrast to the analogical reasoning task applied in Sassenberg et al. (2011), the children in our study were exposed to tasks that did not focus exclusively on spatial information. We thus suggest that the relation between *Gf* and observer viewpoint gestures is a task-specific phenomenon. Furthermore, because current literature shows relations between *Gf* and observer viewpoint gestures only for adults who explain their strategies for solving an analogical reasoning task, it still needs to be investigated whether this relation also applies for children when they explain their solving strategies in an analogical reasoning task.

Conclusion

Our study demonstrates strong associations between visual processing (*Gv)* abilities and the rate of observer viewpoint gestures in young children at the age of 4, indicating that observer viewpoint gestures arise from simulated visual mental imagery. These findings contribute to the growing evidence that in addition to many exogenous factors, the form and the functions of gestures originate from endogenous factors such as cognitive processes. The novelty of our research resides in relating children's cognitive abilities to a specific viewpoint in gesture; and it delivers first empirical evidence on the relations between visual mental images and observer viewpoint gestures assumed in the GSA framework. According to our results, the higher a child's *Gv* abilities, the more she or he will engage with simulated visual imagery in all three tasks. Because we studied children and found this effect in all three tasks with the different cognitive demands they impose, we argue that the relation between *Gv* abilities and observer viewpoint gestures is a general and not a task-specific phenomenon. Whereas there is strong evidence for this relation, it remains an open question whether children with higher *Gv* abilities are more likely to perform observer viewpoint gestures for aspects of events that can be realized with character or observer viewpoint gestures, or whether children with higher *Gv* talk more about spatial aspects of events that primarily elicit observer viewpoint gestures.

References

Abramov, O., Kern, F., Koutalidis, S., Mertens, U., Rohlfing, K., & Kopp, S. (2021). The Relation Between Cognitive Abilities and the Distribution of Semantic Features Across Speech and Gesture in 4-year-olds. *Cognitive Science*, *45*(7), e13012. https://doi.org/10.1111/cogs.13012

Alibali, M. W., & Kita, S. (2010). Gesture highlights perceptually present information for speakers. *Gesture*, *10*(1), 3–28. https://doi.org/10.1075/gest.10.1.02ali

Austin, E. E., & Sweller, N. (2014). Presentation and production: The role of gesture in spatial communication. *Journal of Experimental Child Psychology*, *122*, 92–103. https://doi.org/10.1016/j.jecp.2013.12.008

Bergmann, K., & Kopp, S. (2012). Gestural alignment in natural dialogue. In *Proceedings of the Annual Meeting of the Cognitive Science Society 34*(34), UC Merced, 1326–1331

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture and mental representation in humans and non-human primates. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1585), 129–143. https://doi.org/10.1098/rstb.2011.0162

Cartmill, E. A., Rissman, L., Novack, M., & Goldin-Meadow, S. (2017). The development of iconicity in children's co-speech gesture and homesign. *LIA*, *8*(1), 42–68. https://doi.org/10.1075/lia.8.1.03car

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*(4), 706–723. https://doi.org/10.1037/a0013157

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology. General*, *140*(1), 102–116. https://doi.org/10.1037/a0021790

Church, R. B., Kelly, S. D., & Lynch, K. (2000). Immediate memory for mismatched speech and representational gesture across development. *Journal of Nonverbal Behavior*, *24*(2), 151–174. https://doi.org/10.1023/A:1006610013873

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cook, S. W., Yip, T. K. Y., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594–610. https://doi.org/10.1080/01690965.2011.567074

Debreslioska, S., & Gullberg, M. (2019). Discourse Reference Is Bimodal: How Information Status in Speech Interacts with Presence and Viewpoint of Gestures. *Discourse Processes*, *56*(1), 41–60. https://doi.org/10.1080/0163853X.2017.1351909

Demir, Ö. E., Levine, S. C., & Goldin-Meadow, S. (2015). A tale of two hands: Children's early gesture use in narrative production predicts later narrative structure in speech. *Journal of Child Language*, *42*(3), 662–681. https://doi.org/10.1017/S0305000914000415

Doskocilová, H., Miler, Z., and Jähn, K.-H. (1998/2013). ˇ Der Maulwurf und der Grüne Stern [The Mole and the Green Star]. Leipzig: Leiv

Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology*, *42*(6), 1259–1268. https://doi.org/10.1037/0012-1649.42.6.1259

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Encyclopedia of special education: A reference* *for the education of children, adolescents, and adults with disabilities and other exceptional individuals*. Hoboken, New Jersey: Wiley. https://doi.org/10.1002/9781118660584.ese0431

Frederiksen, A. T. (2017). Separating viewpoint from mode of representation in iconic co-speech gestures: insights from Danish narratives. *Language and Cognition*, *9*(4), 677–708. https://doi.org/10.1017/langcog.2016.35

Gelman, A., & Park, D. K. (2009). Splitting a Predictor at the Upper Quarter or Third and the Lower Quarter or Third. *The American Statistician*, *63*(1), 1–8. https://doi.org/10.1198/tast.2009.0001

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture*, *7*(1), 73–95. https://doi.org/10.1075/gest.7.1.05hos

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A. B., & Alibali, M. W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, *26*(3), 721–752. https://doi.org/10.3758/s13423-018-1548-0

Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language*, *82*(1), 22–29. https://doi.org/10.1016/S0093-934X(02)00009-3

Hostetter, A. B., & Skirving, C. J. (2011). The Effect of Visual vs. Verbal Stimuli on Gesture Production. *Journal of Nonverbal Behavior*, *35*(3), 205–223. https://doi.org/10.1007/s10919-011-0109-2

Kern, F. (2020). Interactional and multimodal resources in children's game explanations. *Research on Children and Social Interaction*, *4*(1), 7–27.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Labov, W., & Waletzky, J. (1973). Erzählanalyse: mündliche Versionen persönlicher Erfahrung. In J. Ihwe (Ed), Literaturwissenschaft und Linguistik. Eine Auswahl Texte zur Theorie der Literaturwissenschaft, Band 2, (pp. 78–126). Hamburg: Athenäum Fischer Taschenbuch Verlag.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Laros, J. A., & Tellegen, P. J. (2015). *Construction and validation of the SON-R 5 1/2-17, the Snijders-Oomen non-verbal intelligence test*. s.n: s.n. Retrieved from https://www.rug.nl/research/portal/files/3344765/LarosTellegen.PDF

Lohman, D. F. (1988). *Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, Vol. 4 (p. 181–248). Lawrence Erlbaum Associates, Inc.*

Lohman, D. F. (1994). Spatial ability. In R. J. Sternberg (Ed.), Encyclopedia of human intelligence (pp. 1000–1007). New York: Macmillan.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 9, 111–151.

McNeill (1992). Hand and Mind: What Gestures Reveal about Thought. Chicago: The University of Chicago Press.

McNeill D. (1992). *Hand and mind: What gestures reveal about thought.*

Mickley, M., & Renner, G. (2010). Intelligenztheorie für die Praxis: Auswahl, Anwendung und Interpretation deutschsprachiger Testverfahren für Kinder und Jugendliche auf Grundlage der CHCTheorie. Klinische Diagnostik und Evaluation, 3, 447--466. *Klinische Diagnostik Und Evaluation*, *3*, 447--466.

Mickley, M., & Renner, G. (2019). Auswahl, Anwendung und Interpretation deutschsprachiger Intelligenztests für Kinder und Jugendliche auf Grundlage der CHC-Theorie: Update, Erweiterung und kritische Bewertung [Selection, Use, and Interpretation of German Intelligence Tests for Children and Adolescents Based on CHC-theory: Update, Extension, and Critical Discussion]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *68*(4), 323–343. https://doi.org/10.13109/prkk.2019.68.4.323

Özer, D., & Göksun, T. (2020). Gesture Use and Processing: A Review on Individual Differences in Cognitive Resources. *Frontiers in Psychology*, *11*, 573555. https://doi.org/10.3389/fpsyg.2020.573555

Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, *25*(5), 650–668. https://doi.org/10.1080/01690960903424248

Perry, M., Church, R.B., & Goldin-Meadow, S. (1992). Is gesture-speech mismatch a general index of transitional knowledge? *Cognitive Development*, *7*(1), 109–122. https://doi.org/10.1016/0885-2014(92)90007-E

Quasthoff, U., Heller, V., & Morek, M. (2017). On the sequential organization and genre-orientation of discourse units in interaction: An analytic framework. Discourse Studies, 19(1), 84–110.

Raven, J. C. (1958). Advanced progressive matrices. London: Lewis.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Sassenberg, U., Foth, M., Wartenburger, I., & van der Meer, E. (2011). Show your hands — Are you really clever? Reasoning, gesture production, and intelligence. *Linguistics*, *49*(1), 841. https://doi.org/10.1515/ling.2011.003

Sassenberg, U., & van der Meer, E. (2010). Do we really gesture more when it is more difficult? *Cognitive Science*, *34*(4), 643–664. https://doi.org/10.1111/j.1551-6709.2010.01101.x

Schretlen, D., Pearlson, G. D., Anthony, J. C., Aylward, E. H., Augustine, A. M., Davis, A., & Barta, P. (2000). Elucidating the contributions of processing speed, executive ability, and frontal lobe volume to normal age-related differences in fluid intelligence. *Journal of the International Neuropsychological Society*, *6*(1), 52–61. https://doi.org/10.1017/S1355617700611062

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., . . . Uhmann S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion,, *10*, 353–402.

Smithson, L., & Nicoladis, E. (2014). Lending a Hand to Imagery? The Impact of Visuospatial Working Memory Interference Upon Iconic Gesture Production in a Narrative Task. *Journal of Nonverbal Behavior*, *38*(2), 247–258. https://doi.org/10.1007/s10919-014-0176-2

Stec, K. (2012). Meaningful shifts. *Gesture*, *12*(3), 327–360. https://doi.org/10.1075/gest.12.3.03ste

Stites, L. J., & Özçalışkan, Ş. (2017). Who Did What to Whom? Children Track Story Referents First in Gesture. *Journal of Psycholinguistic Research*, *46*(4), 1019–1032. https://doi.org/10.1007/s10936-017-9476-0

Tellegen, Laros, & Petermann (2007). SON-R 2½- 7 Non-verbaler Intelligenztest. Testmanual mit deutscher Normierung und Validierung. Göttingen: Hogrefe.

Vernon, P. (1983). Speed of information processing and general intelligence. *Intelligence*, *7*(1), 53–70. https://doi.org/10.1016/0160-2896(83)90006-5

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

SUPPLEMENTARY ITEMS

Table 1: Script for The Game in The Communicative Task of Explanation

Instruction	**Translation**
Gleich wird [name of experimenter] Ihrem Kind ein Spiel zeigen. Wenn Sie wieder in den Raum hineinkommen, setzen Sie sich bitte auf ein Kissen gegenüber Ihrem Kind. Die Plätze sind fest vorgesehen, damit die Kamera Ihr Kind gut erfasst.	[name of experimenter] is about to show your child a game. When you come back into the room, please sit on a cushion opposite your child. The seats are fixed so that the camera can capture your child well.
Wenn Sie sitzen, fragen Sie dann bitte Ihr Kind, was das für ein Spiel war und ob es Ihnen den Ablauf erklären kann. Das Spiel können Sie mit Ihrem Kind gern am Ende der Sitzung spielen. (z.B. „Erklär' mir das Spiel, damit wir es gleich spielen können!“)	When you are seated, please ask your child what kind of game it was and if she or he can explain the procedure to you. You are welcome to play the game with your child at the end of the session. (e.g., “Explain the game so we can play it right away!”)
Falls Ihr Kind keinen Erzählanfang findet, können Sie ihm gern mit folgenden Fragen helfen (dies sind nur Vorschläge, sie brauchen sich diese also nicht alle merken):	If your child cannot find a starting point for the story, you can help her or him with the following questions (these are only suggestions, so you do not need to remember them all):
Worum geht es in dem Spiel?	What is the game about?
Gab es in dem Spiel eine Spielfigur?	Was there a character in the game?
Was war auf dem Würfel?	What was on the dice?
Wann darf der Junge fliegen?	

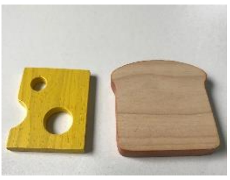
Wann habe ich gewonnen?	When is the boy allowed to fly? When do I win?

Table 2: The Board Game (in The Communicative Task *Explanation*)

	Stimuli used	
1	Self-made jigsaw-puzzle (20 cm x 35 cm) included jigsaw pieces: - Triangle (as a part of the tower) - Moon - Star - Circle (as a part of the house) - Rectangle (as a part of the house)	
2	Figure (1,5 cm x 4 cm) from Playmobil®: This was the protagonist of the game, who could “fly” through the shapes.	
3	Wooden dice with yellow shapes on it. The shapes accorded with the shapes available in the jigsaw-puzzle: - Triangle - Moon - Star - Circle - Rectangle - Blank	

Table 3: The Five Events That Were Performed by The Hand Puppet

in The Communicative Task *Illustration With Examples*

	Stimuli	**Performed event**
	Hand puppet 	The dog performed all the following events:
1	Plastic plate and a plastic spoon (both children's kitchenware) 	*Verbal behavior:* "Ja, ich habe gesehen, wie die Menschen Löffel halten, nämlich so!" [I have seen how the people hold a spoon, like this!] *Nonverbal behavior:* Dog holds the spoon on the wrong side and ladles something with the handle
2	Plastic bottle (small size, usual type) 	*Verbal behavior:* "Und ich weiß, wie man aus einer Flasche trinkt, nämlich so!" [And I know how to drink from a bottle, like this!] *Nonverbal behavior:* Dog turns the bottle visibly upside down and drinks from the bottom
3	Slice of cheese and a piece of bread (children's playware) 	*Verbal behavior:* "Und ich weiß, wie ich Käse auf mein Brot tue, nämlich so!" [And I know how to put a slice of chees on my bread, like this!] *Nonverbal behavior:* Dog puts the slice on the bread, so it barely touches it (almost beside the bread)
4	Pot from children's kitchenware and a saltshaker (regular size) 	*Verbal behavior:* "Und ich weiß, wie ich Salz in meine Suppe tun kann, nämlich so!" [And I know how to put salt in my soup, like this!] *Nonverbal behavior:* Dog put the salt on the table first and tries to take a corn of salt and to put it then in the pot.

5	Cup (regular size) and a tea bag with a string 	*Verbal behavior:* “Und ich weiß, wie man einen Tee macht, nämlich so!“ [And I know how prepare a tea, like this!] *Nonverbal behavior:* Dog takes a tea bag and throws it entirely in the cup (for a tea bag with a string, a piece of the string has to be left out of the cup)