

## Language and Cognitive Processes

Publication details, including instructions for authors and  
subscription information:

<http://www.tandfonline.com/loi/plcp20>

### Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure

Fey Parrill<sup>a</sup>

<sup>a</sup> Case Western Reserve University, Cleveland, OH, USA

Available online: 20 Jan 2010

To cite this article: Fey Parrill (2010): Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure, *Language and Cognitive Processes*, 25:5, 650-668

To link to this article: <http://dx.doi.org/10.1080/01690960903424248>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure

Fey Parrill

*Case Western Reserve University, Cleveland, OH, USA*

We examine a corpus of narrative data to determine which types of events evoke character viewpoint gestures, and which evoke observer viewpoint gestures. We consider early claims made by McNeill (1992) that character viewpoint tends to occur with transitive utterances and utterances that are causally central to the narrative. We argue that the structure of the event itself must be taken into account: there are some events that cannot plausibly evoke both types of gesture. We show that linguistic structure (transitivity), event structure (visuo-spatial and motoric properties), and discourse structure all play a role. We apply these findings to a recent model of embodied language production, the Gestures as Simulated Action framework.

**Keywords:** Embodiment; Event structure; Gesture; Perspective; Viewpoint.

Imagine that you are describing the scene shown in Figure 1, in which a cartoon skunk hops across a room. In describing this event, it is very likely that you will produce hand gestures, even without being consciously aware of having done so (Goldin-Meadow, 2003; McNeill, 1992, 2005). There are several possibilities for how the gestures that accompany your speech might encode information about the scene. For example, you might trace the path the character took with your hand, as shown in Figure 2. Such gestures (in

---

Correspondence should be addressed to Fey Parrill, Case Western Reserve University, Crawford Hall, 612B, 10900 Euclid Ave, Cleveland, OH 44106-7063, USA. E-mail: fey.parrill@case.edu

The author is grateful to Mark Turner, Sotaro Kita, and three anonymous reviewers for comments and suggestions regarding data analysis. The issue of gestures that represent visual imagery requiring a process of transduction was first raised for the author by Daniel Casasanto. She also thanks the research assistants in the Gesture and Cognition Lab who worked on data collection and analysis, in particular, Amanda Dewitt, Rochelle Hudson, Huston Hoburg, Caitlin Dawson, and Noel Hanzel.

---

© 2010 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business  
<http://www.psypress.com/lcp> DOI: 10.1080/01690960903424248



**Figure 1.** Stimulus being described.

which the hand represents the character and the trajectory traced by the hand represents the character's path) have been called *observer* viewpoint gestures (McNeill, 1992). The action is depicted as though you are observing it from afar. Alternately, you might show the action of the character's hands with your own hands, as in Figure 3. Such gestures (wherein a speaker uses her own body to show how the character acted) have been called *character* viewpoint gestures (McNeill, 1992).

Researchers who study multimodal language have long been aware of this phenomenon in gesture, but very little is known about what makes a narrator



**Figure 2.** Observer viewpoint gesture.



**Figure 3.** Character viewpoint gesture.

choose one or the other representation. McNeill (1992) has suggested that the transitivity of the accompanying utterance and the centrality of the event being described may be determining factors. These claims need to be verified, however: it seems likely that some events, because of their spatial-motoric properties, cannot plausibly evoke both types of gesture. Furthermore, viewpoint in gesture has important implications for recent *embodied* theories of language (Gibbs, 2006; Lakoff & Johnson, 1999). While these theories will be discussed in greater detail below, many proponents claim that using language involves creating a mental simulation of the content being communicated (Glenberg & Kaschak, 2002; Lakoff & Johnson, 1999; Zwaan, 1999). Gestures represent a *physical* simulation of the representations that give rise to language. They can therefore provide a new source of support for embodiment, as Hostetter and Alibali have recently argued (2008). Gestures also suggest that there are differences in the perspective taken during simulation, as others have argued (Bergen, Lindsay, Matlock, & Narayan, 2007; Lozano, Martin Hard, & Tversky, 2007). A character viewpoint gesture may represent simulation from a first-person point of view, while an observer viewpoint gesture may represent simulation from a third-person point of view. There are reasons to believe taking different points of view uses different neural mechanisms (Ruby & Decety, 2001). Using PET, Ruby and Decety determined that different cortical circuits were involved when participants were asked to mentally simulate a sentence describing an action from the participant's point of view (e.g., *you are stapling a sheet of paper*) as compared with a third-person point of view (e.g., *I am stapling a*

*sheet of paper*). In short, for all these reasons, gestural viewpoint represents an essential topic for investigation.

In this study, we examine a corpus of narrative data to determine which types of event evoke character viewpoint gestures and which evoke observer viewpoint gestures. We argue that linguistic structure (transitivity), event structure, and discourse structure all play a role. The structure of the paper is as follows. We first provide a brief overview of viewpoint in gesture and present McNeill's claims regarding transitivity and centrality. We then discuss embodied theories of language, with a focus on Hostetter and Alibali's Gestures as Simulated Action framework. We next provide evidence that multiple factors determine whether a participant is likely to adopt a particular gestural viewpoint. We end by returning to the significance of viewpoint in gesture for theories of embodied cognition. We discuss our findings in relation to the predictions made by the Gestures as Simulated Action framework.

## VIEWPOINT IN CO-SPEECH GESTURE

Viewpoint is defined for this study as the subject's perspective on an event or scene (Chafe, 1976; DeLancey, 1981). (But see Parrill, 2009b, for discussion of the complexities involved in defining viewpoint.) In gesture, viewpoint manifests itself in a number of ways. For the current purposes, the most important distinction is between an internal, first-person point of view (character viewpoint, or C-VPT) and a more external, third-person point of view (observer viewpoint, or O-VPT). Many researchers have commented on the distinction between character and observer viewpoint in passing, but little effort has been made to provide an explanation for which circumstances evoke one or the other. The most systematic accounts are found in research on route descriptions, although this work uses different terminology to describe the same essential alternation. Tversky and colleagues, for example, have shown that a more internal point of view tends to be adopted when describing spaces that have a single scale, while an external point of view is used when the space being described varies in scale (Bryant & Tversky, 1999; Emmorey, Tversky, & Taylor, 2000; Taylor & Tversky, 1996). It is unclear, however, whether these findings can be extended to narrative language. For narrative language, only a few generalisations have been made. Parrill (2009b) has shown that common ground (in this case, knowing a partner has access to information) shared between speech participants can impact gestural viewpoint. Beattie and Shovelton (2002) have shown that C-VPT conveys more information to interlocutors. And, in assessing the circumstances that *elicit* C-VPT, McNeill (1992) has described two basic patterns.

First, McNeill notes that C-VPT tends to occur with transitive events (1992, p. 119). Second, McNeill notes that C-VPT tends to occur with events that are causally central to the discourse (1992, p. 120). These claims need to be reassessed for several reasons. To begin with, they rely on approximately four narratives of a single stimulus (the bulk of the data being available only as an unpublished conference paper: Church, Baker, Bunnag, & Whitmore, 1989). In addition, these claims appear to assume that both points of view are equally possible for any given event, provided transitivity is held constant. As we will show, in order to come to an accurate understanding of the constraints on gestural viewpoint, the nature of the event being described must be taken into account. In addition, if using language does involve creating a mental simulation – as proponents of embodied cognition believe – it becomes particularly important to consider the properties of the event being simulated. We turn now to theories of embodied language.

## VIEWPOINT AND EMBODIMENT

Embodiment refers generally to the idea that cognitive constructs are shaped by our interactions with the world using our human bodies (Gibbs, 2006). (For a concise comparison of embodiment with other approaches to representation, see Markman & Dietrich, 2000.) Semantic constructs, such as the meanings of words, are claimed to involve modality-specific representations, rather than being purely symbolic (Glenberg & Kaschak, 2002; Zwaan, 1999, Zwaan, Stanfield, & Yaxley, 2002). For example, one's understanding of the word *hammer* might rely on a mental image of the object and a motor program for how one grips the object. Embodied theories of language claim that more complex instances of language use, such as comprehending and producing sentences, involve mentally simulating the events being talked about. A simulation is simply a partial reconstruction of an event, but is generally understood as involving the generation of mental images and the activation of motor programs. Simulations are thought to use some of the same neural substrates as visual perception or motor action (see Barsalou, 2008 for a recent summary). When producing or understanding a sentence like *Sarah opened the drawer*, for example, one might generate a mental image of the scene and activate a motor program having to do with motion of the arm towards the body. Some support for these claims exists in the domain of sentence comprehension. Psycholinguistic research (Glenberg & Kaschak, 2002; Kaschak & Glenberg, 2000) has shown that participants are slower to respond when required to produce an action that is incompatible with a motion implied by the meaning of a sentence (e.g., a motion of the arm away from the body would be incompatible with *Sarah opened the drawer*). Glenberg and Kaschak explain this pattern by suggesting

that when sentences involve motor actions, participants automatically activate motor programs during comprehension. Similar effects can be obtained for imagery: when participants are asked to make a judgement about a picture that does not match a sentence they have just read, they are slower to respond (Zwaan, Stanfield, & Yaxley, 2002). Zwaan and colleagues suggest this is because a mental image is automatically generated when reading a sentence.

While much of the work on embodied cognition has focused on language (and typically language comprehension), the basic claims of the framework are quite broad. Gibbs (2006) provides a comprehensive picture of embodiment in cognitive domains ranging from perception to emotion. Lakoff and Johnson (1999) describe the implications of embodiment for philosophy, linguistics, and ethics. Embodiment is also central to many recent developments in artificial intelligence (AI) (Clark, 1997), and researchers are beginning to provide specific neuroscientific details about the ways embodied experiences might have implications for AI (Downing, 2007).

Gestures can offer support for embodied frameworks, because they involve not just mental, but *physical* simulations of the representations that give rise to language. Indeed, Hostetter and Alibali (2008) have proposed a model that attempts to unify these domains of research. According to their Gestures as Simulated Action (GSA) model, ‘... gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery’ (p. 502). That is, cognition in the context of language use involves visual and motor simulation, and gestures reflect this simulation.

How exactly can a gesture be seen as a simulation? According to the GSA model, during language production mental images and motor programs are activated as part of the process of simulation. These simulations result in motor activation, which may manifest itself as gesture. (Details about the circumstances under which a person will gesture must be passed over here, but can be found in Hostetter and Alibali, 2008.) For example, if talking about an agent performing an action (say, *Sarah passed the salt*), a motor program for arm motion away from the body with a certain hand shape will be activated. This might manifest itself as a character viewpoint gesture. A mental image of a person passing an object will also be generated. Since the sentence is describing the action of a third person, the perspective taken in simulation may be that of an observer, rather than an agent. In that case, the visual trajectory from the mental image may be re-represented as a motor action tracing that trajectory. The speaker may thus produce an observer viewpoint gesture, moving laterally.

The GSA model focuses on integrating what is known about simulation with what is known about gesture, and as a result the authors do not discuss

any aspect of gesture with great detail. However, Hostetter and Alibali do suggest that motor simulation might be more likely to evoke character viewpoint gestures, while simulation of visual imagery might be more likely to evoke observer viewpoint gestures. We agree with this general claim, though we question the assumption that visual imagery and motor imagery are wholly separable in gesture. First, some gestures represent both motoric and imagistic features of a mental representation. (For example, dual viewpoint gestures, which simultaneously encode both a character's and an observer's point of view: McNeill, 1992, Parrill, 2009a.) Second, gestures that represent visual imagery must involve some process of transduction or re-representation, because the output of the system is a motor action. The question of how mental images are converted into motor actions is not resolved within the GSA model. Indeed, this is a larger issue in cognitive science (see, for example, Paivio, 1986). In the final section of the paper, we return to the claims of the GSA model in the domain of viewpoint.

## OUR EXTENSIONS

This study extends research on multimodal language in three ways. First, we test McNeill's claims about viewpoint with a much larger dataset of narrative language (23 narrations of three different stimuli, or 69 narrations). Second, we show that while linguistic structure and discourse structure play a role in determining the selection of gestural viewpoint, *event structure* is also crucial. By event structure we mean spatial, imagistic, and motion properties of an event, including features like the affordances (Gibson, 1966) of objects involved. We show that event structure exerts more influence in some cases, while discourse factors may exert more influence in others. Finally, we apply these findings to the Gestures as Simulated Action framework, providing data that support some of the framework's predictions.

## DATA AND ANALYSIS

### Participants and method

Forty-six Case Western Reserve University students (31 women) participated in the study for payment. All were native speakers of American English. Participants came to the experiment room with a friend. (People tend to gesture more naturally when talking to a friend: Parrill, in press.) These pairs decided who would serve as the narrator and who would serve as the listener. Following informed consent, narrators watched three 30-second video clips in random order. The clips were excerpts from cartoons, and are described in detail in Appendix A. After watching each clip, the narrator described it to



his or her partner. These narrations were videotaped. Following the study, listeners were given a series of comprehension questions asking about basic properties of the stimuli (e.g., *In the skunk cartoon, were any of the characters human?*). The purpose of this quiz was simply to motivate narrators to describe the events in detail.

## Coding

We carried out two kinds of coding: coding that focused on the events of the stimuli, and coding that focused on the speech and gesture produced by participants.

*Event coding.* Video stimuli were first divided into events. Two independent coders divided each clip into a series of events based on changes of viewing angle and actions of characters. Agreement was 75%, and any event not listed by both coders was excluded, to produce a final set of events for the analyses. Next, each event was coded by two independent coders as either central to the discourse or peripheral to the discourse, using criteria described in Stein and Glen (1979). Agreement was 84%, and discrepancies were resolved through discussion. Details on events and discourse centrality can be found in Appendix A. It should be noted that this coding procedure does not correspond perfectly to that of Church and colleagues, as described in McNeill (1992). Church and colleagues assigned discourse centrality to utterances produced by narrators rather than to events in their stimulus. We achieve this same end with a later step of coding, described below. We assigned discourse status to events in the stimuli in order to maintain consistency across a much larger data set, and to allow additional patterns to be observed. However, in order to ensure that any differences between our results and those of Church and colleagues are not due to this difference in method, we carried out an additional analysis on a subset of the data. A randomly selected sample of seven narrations of each stimulus (or 30% of the total dataset) was analysed according to the Church et al. method as well. That is, discourse centrality was coded for each utterance. Two independent coders carried out this analysis, and agreed on coding for 91% of utterances.

*Gesture coding.* All gestures produced by narrators were transcribed and sorted into four categories: character viewpoint, observer viewpoint, dual viewpoint (simultaneously expressing character and observer viewpoint), or no viewpoint. This latter category includes metanarrative gestures with metaphoric content, purely rhythmic gestures, deictic gestures, and certain types of iconic gestures that describe static aspects of a scene (e.g., two hands tracing the outline of an object). Three independent coders carried out this analysis for the entire dataset. Agreement was 70%, and discrepancies were

resolved through discussion. Analysis will focus only on C- and O-VPT gestures.

*Speech coding.* All speech produced by narrators was transcribed. Each utterance that was accompanied by a C- or O-VPT gesture was marked as transitive, intransitive, or neither (used for utterance fragments or other non-predicative elements, such as *then like bam*) by two independent coders for the entire dataset. Agreement was 87%, and discrepancies were resolved through discussion.

*Matching speech to events.* Finally, two independent coders matched each utterance accompanied by a C- or O-VPT gesture to an event from the stimulus, for the entire dataset. Agreement was 81%. Utterances that did not relate to established events were excluded from analysis. These tended to be meta-narrative commentary (e.g., *he was kind of creepy*).

RESULTS

The initial corpus contained 892 gestures. Relative frequencies of the different types are as follows: C-VPT: 27%; O-VPT: 29%; D-VPT: 0.4%; N-VPT: 42%. Because we are interested solely in C- and O-VPT gestures, analysis will focus on this subset (a total of 506 gestures, of which 46% were C-VPT).

C-VPT and transitivity

Table 1 shows the frequency of each type of utterance according to the viewpoint of the gesture it occurred with. Transitive utterances were less frequent than intransitive utterances in these data:  $\chi^2(1) = 30.58, p < .0001$ .

Different participants produced different numbers of gestures, so we carried out an analysis examining the proportion of C-VPT and O-VPT gestures each participant produced according to utterance type. As shown in Figure 4, C-VPT utterances were more frequent with transitive utterances than with intransitive utterances ( $t(22) = -5.61, p < .0001$ ).

TABLE 1  
Viewpoint and utterance transitivity

	<i>Transitive utterance</i>	<i>Intransitive utterance</i>	<i>Neither</i>	<i>Total</i>
C-VPT	130 (56%)	78 (33%)	25 (11%)	233
O-VPT	45 (16%)	218 (80%)	10 (4%)	273
Total	175	296	35	506

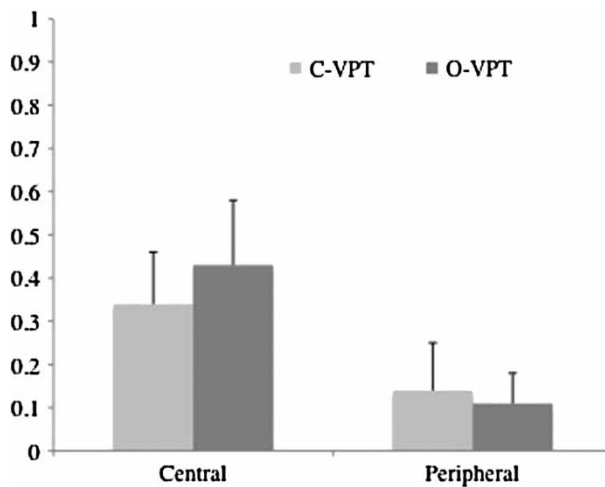


Figure 4. Proportion of each gesture type according to sentence type.

### C-VPT and discourse centrality

More gestures were produced for discourse central events, but this is due to the fact that central events are more frequent. The first two rows of Table 2 show the mean proportion of each gesture type according to the centrality of the event the accompanying utterance matched. Participants produced more C-VPT gestures for central events as compared with peripheral events (paired  $t$ -test:  $t(22) = 5.81$ ,  $p < .0001$ ). They also produced more O-VPT gestures for central events as compared with peripheral events (paired  $t$ -test:  $t(22) = 8.85$ ,  $p < .0001$ ). The comparison of primary interest, however, is between different types of gestures produced for central events alone. Do central events actually evoke more C-VPT? A paired  $t$ -test on the mean number of C-VPT versus O-VPT gestures produced for discourse central events showed no significant difference ( $t(22) = -1.47$ ,  $p = .15$ ).

When viewpoint is examined in relation to the centrality of an utterance (rather than the event it describes), the same pattern emerges. The bottom

TABLE 2  
Viewpoint and discourse centrality of event: mean proportion of each gesture type

	Discourse central	Discourse peripheral
By event: C-VPT	0.34 ( <i>SD</i> 0.12)	0.14 ( <i>SD</i> 0.11)
By event: O-VPT	0.43 ( <i>SD</i> 0.15)	0.11 ( <i>SD</i> 0.07)
By utterance: C-VPT	0.34 ( <i>SD</i> 0.29)	0.14 ( <i>SD</i> 0.20)
By utterance: O-VPT	0.36 ( <i>SD</i> 0.20)	0.06 ( <i>SD</i> 0.12)

two rows of Table 2 show (for the subset of the data described above) the mean proportion of each gesture type according to the centrality of the accompanying utterance. Again, a paired *t*-test on the mean number of C-VPT versus O-VPT gestures produced for discourse central events showed no significant difference ( $t(20) = -.21, p = .83$ ).

### Viewpoint and event structure

When gesture was examined in relation to stimulus events, several additional patterns emerged. These patterns suggest that event structure imposes constraints on the expression of viewpoint in gesture. Our analyses indicated that some events evoke exclusively C-VPT, some evoke exclusively O-VPT, and some evoke both.

*C-VPT only.* Events in which handling (the use of a character's hand in accomplishing some task) is prominent evoked exclusively C-VPT. For example, in one stimulus, a character rides a bus, holding and reading a newspaper. This event is typically described with a transitive utterance (usually *he reads a newspaper*), but transitivity may not be the best explanation for the appearance of C-VPT. A better explanation may be that there is no obvious O-VPT representation for reading a newspaper. Table 3 shows the events that

TABLE 3  
Events evoking only C-VPT gestures

Event #	Event description	# Gestures	Type
K2	Woman complains about dog (taps foot)	10	Affect
K4	Kitten is on dog's back	9	Torso
K5	Dog puts kitten on shelf	19	Handling
K6	Dog covers kitten with bowl	12	Handling
K7	Dog agrees w/woman (nods)	6	Affect
K11	Dog is laughing	2	Affect
K13	Dog is alarmed	3	Affect
R2	Rabbit throws ball	12	Handling
R3	Rabbit looks alarmed	4	Affect
R7	Rabbit reads paper	9	Handling
R13	Rabbit hooks self onto flagpole	8	Handling
R16	Rabbit throws mitt up	18	Handling
R17	Mitt catches ball	12	Handling
R18	Mitt falls back onto hand	18	Handling
S1	Skunk holds cat and kisses	11	Handling
S3	Skunk shrugs	2	Torso
S10	Cat closes door	12	Handling
S11	Cat leans against door, sighs with relief	13	Torso

evoked exclusively C-VPT (18 total). These events can be sorted into four categories: those that involved handling (ten total), those that involved some kind of emotional state or affect (e.g., alarm, laughter: five total), and those that involved some use of the torso (e.g., shrugging, leaning) that cannot be readily depicted from an observer's point of view (three total). Some of these events *do* have plausible O-VPT variants. For example, while participants tended to point to their own bodies (C-VPT) when describing event K4 (Kitten is on dog's back), it is possible to use one's hand to represent the dog's body and point to a location on the hand (O-VPT). Despite this possibility, participants never took an observer's point of view on this event. In short, some events are gestured only from a character's point of view, regardless of how central the event may be, or whether it involves a direct object. In such cases, properties of the event itself (the fact that it involves a character's hand/torso, or affectual state) are the best explanation for viewpoint choice in gesture.

*O-VPT only.* Events that involved trajectories tended to evoke exclusively O-VPT gestures. For example, an event in which a character runs out of a baseball stadium evoked only O-VPT. Events with trajectories are (usually) subject-verb-prepositional phrase, thus intransitive. However, the fact that a trajectory can be more easily gestured from an observer's point of view seems to be driving this pattern, rather than transitivity. All 18 of the events that evoked exclusively O-VPT involved a trajectory of some sort (Table 4 shows these events). However, the structure of these events does not preclude the production of C-VPT gesture: Some of these events *can* potentially be gestured from a character's point of view. For example, a narrator might move her arms as though running when describing the character's transit out of a baseball stadium, thus focusing on an internal aspect of the action, rather than the trajectory. Furthermore, some of the events that evoked only C-VPT also involve trajectories (e.g., R16), thus trajectory alone cannot explain this pattern. As with the events above, it seems that event structure strongly predisposed narrators to adopt a particular point of view, though it may not completely constrain the choice.

*Both C-VPT and O-VPT.* In each stimulus, certain events evoked C and O-VPT gestures with roughly equal frequency. In such cases, discourse factors (such as the narrator's focus, the status of the information as given or new, the structure of the narrative) appear to determine which point of view is taken in gesture, rather than the structure of the event itself. Table 5 shows these events. Of particular interest is event S4. S4 has exactly the same properties as event S6 above, yet this latter event evokes only O-VPT. The difference between the two is presumably one of narrative structure, and therefore discourse relevance. S4 is the first event in which the character exhibits a marked and amusing hopping motion (shown in Figure 1 above).

TABLE 4  
Events evoking only O-VPT gestures

<i>Event #</i>	<i>Description</i>	<i># Gestures</i>
K1	Dog walks up to house	2
K3	Dog comes into room	9
K8	Kitten is going into room and dog follows	2
K14	Ball (with kitten on top) rolls towards woman	19
K15	Ball (with kitten on top) rolls into woman's leg	22
R4	Ball flies out of stadium	17
R5	Rabbit chases ball out of stadium	19
R6	Rabbit takes bus to building	8
R8, 12, 15	Ball flies overhead	5
R9	Bus arrives at building, rabbit gets off	6
R10	Rabbit takes elevator to top of building	17
R11	Rabbit gets off elevator, runs across roof	3
S2	Cat escapes	17
S5	Cat runs, banking turn on wall	24
S6	Skunk follows, hopping	9
S7	Cat goes up stairs	13
S8	Skunk follows up stairs	13
S9	Cat goes into room	8

TABLE 5  
Events evoking both C- and O-VPT gestures

<i>Event #</i>	<i>Description</i>	<i># Gestures</i>	<i>Proportion CVPT</i>
K10	Kitten is playing w/ball (tumbling on and off)	12	0.58
K12	Kitten scrambles on ball, ball starts to roll	12	0.33
R14	Rabbit pulls self up flagpole	24	0.58
S4	Skunk hops after cat	24	0.61

The first mention of this action, therefore, tends to evoke both focus on the internal structure of the character's action (C-VPT) and the trajectory (O-VPT), while later descriptions of the event evoke focus on the trajectory alone. Discourse centrality does not appear to be relevant, inasmuch as while most of these events were coded as discourse central (the exception being K10), not all discourse central events evoke this kind of 'free variation'.

## DISCUSSION

This study aimed to answer three specific questions. First, do C-VPT gestures tend to occur with transitive utterances? We find that they do, confirming McNeill's claim. Second, we asked whether C-VPT gestures tend

to occur with events that are causally central to the narrative. Our results suggest that this is *not* the case. While discourse-central events evoked more C-VPT gestures than did discourse-peripheral events, this is due to the fact that central events evoke more gestures *in general*. Our data suggest that McNeill's earlier generalisation (based on the results of Church et al., 1989) may have been the product of a smaller dataset, a failure to account for the imbalance in number of gestures occurring with utterances describing central versus peripheral events, or the nature of the events in the Church et al. study. Because we analysed a subset of our data using their method, this discrepancy is not likely to be due to a methodological difference.

Finally, we asked whether additional patterns emerge as a function of event structure (rather than discourse structure or transitivity). We find that event structure strongly predisposes narrators to adopt a particular viewpoint in many cases. Events that involve a display of affect or a prominent use of the character's hands and torso predisposed narrators to adopt the viewpoint of the character. Events involving trajectories strongly predisposed narrators to adopt the viewpoint of an observer. These properties may offer better explanations for narrator's behaviours than transitivity. However, while event structure seems to exert a strong influence in the majority of cases, we also find that, for certain events, the expression of viewpoint in gesture varies from participant to participant. Such events are too infrequent in our data to permit any broad generalisations, but discourse structure – including the salience or accessibility of a referent or action – appears to play a role. Our data provide an example where the first mention of an event evokes focus on the internal structure of an event for some narrators and on the trajectory for others. The second mention of the same event, however, evokes exclusively focus on trajectory.

A natural question to ask about events that seem to be in 'free variation' is whether narrators can be pushed in one direction or the other. We are currently carrying out a series of studies designed to answer this question, examining factors such as verbal aspect and stimulus modality (text vs. video). More generally, the patterns described here may serve as a starting point for manipulations of viewpoint, and for comparisons between viewpoint in gesture and parallel phenomena in signed languages, namely the device known as *constructed action* (Liddell & Metzger, 1998; Quinto-Pozos, 2007; Quinto-Pozos, Cormier, & Holzrichter, 2006). A recent study comparing American Sign Language and English speech and gesture along a number of dimensions suggests that character viewpoint and constructed action share many features, but also differ in crucial ways (Marentette, Tuck, & Nicoladis, *in review*). Quinto-Pozos and Parrill (2008) are carrying out several studies focusing specifically on viewpoint in the two language systems.

Finally, what significance do our data have for theories of embodied cognition? While both C- and O-VPT gestures suggest a narrator is

simulating an event, these data may indicate that embodiment can vary perspective, with some events being simulated from the point of view of an agent, and others from the point of view of an observer. If so, it appears that the event itself exerts a strong influence in determining how language users will simulate, for the majority of cases. For some cases, however, language users may have a choice in the matter: these cases offer a particularly interesting object for further study.

## VIEWPOINT IN THE GESTURES AS SIMULATED ACTION FRAMEWORK

We would like to end by addressing three specific predictions about viewpoint made by Hostetter and Alibali. First, Hostetter and Alibali suggest that C-VPT gestures may arise from motor simulation, while O-VPT gestures might arise from simulation of visual imagery. As noted above, we believe this is a simplification, but our findings certainly are in accord with this claim. Events in which handling and actions of the character's torso or face were prominent evoked exclusively C-VPT, and these events should involve motor simulation. Events that evoked exclusively O-VPT gestures tended to involve a long trajectory that cannot be simulated at a human scale, and these events should involve simulation of visual imagery, which must then be translated into a motor action. Second, Hostetter and Alibali suggest that events which one has personally experienced might be more likely to evoke C-VPT (p. 504). Data from our lab (Parrill, 2009c) suggest that simply describing an event as though one has experienced it may have the same effect. Participants asked to describe events as though they had performed them produced more C-VPT gestures than those given no instructions. Thus, whether or not one has experienced something, one can simulate it as though one has. Finally, Hostetter and Alibali suggest that modality of input may have an impact on gestural viewpoint. We find that this may *not* be the case. In a study comparing narrations produced after reading a text to narrations produced after watching a video, we find that viewpoint is not affected (Parrill, Bullen, & Hoburg, 2009). (Hostetter and Hopkins, 2002 carried out the same manipulation, but did not explore viewpoint.)

## CONCLUSION

We have shown that event structure (spatial, imagistic, and motion properties of an event) and linguistic structure (transitivity) play a role in determining viewpoint in gesture, for narrative language. We have also shown that discourse factors may exert influence. Finally, we have attempted to apply these findings to a current model of multimodal language, the Gestures as



Simulated Action framework. Our findings support several of the claims made by Hostetter and Alibali, while ongoing work in our lab will provide further tests of their predictions. We believe viewpoint to be central to discussions of embodied cognition. We also believe that gesture offers strong support for such theories, and hope that the data and proposals here will serve as starting points for future research.

Manuscript received March 2009

Revised manuscript received October 2009

First published online January 2010

## REFERENCES

- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology*, 93, 179–192.
- Bergen, B., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, 31(5), 733–764.
- Bryant, D. J., & Tversky, B. (1999). Mental representations of perspective and spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 137–156.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li & S. A. Thompson (Eds.), *Subject and topic* (pp. 27–55). New York: Academic Press.
- Church, R. B., Baker, D., Bunnag, D., & Whitmore, C. (1989). The development of the role of speech and gesture in story narration. *Biennial Meeting of the Society for Research in Child Development*. Kansas City, MO.
- Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.
- DeLancey, S. (1981). An interpretation of split ergativity and related patterns. *Language*, 57(3), 626–657.
- Downing, K. L. (2007). Neuroscientific implications for situated and embodied artificial intelligence. *Connection Science*, 19(1), 75–104.
- Emmorey, K., Tversky, B., & Taylor, H. A. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, 26, 157–180.
- Gibbs, R. (2006). *Embodiment and cognitive science*. Cambridge, UK: Cambridge University Press.
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Prospect Heights, IL: Waveland Press, Inc.
- Glenberg, A., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, 96, 558–565.
- Goldin-Meadow, S. (2003). *Hearing gesture: how our hands help us think*. Cambridge, MA: Belknap Press of Harvard University Press.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin and Review*, 15(3), 495–514.
- Hostetter, A. B., & Hopkins, W. D. (2002). The effect of thought structure on the production of lexical movements. *Brain and Language*, 82, 22–29.
- Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43(3), 508–529.

- 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.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.
- Liddell, S., & Metzger, M. (1998). Gesture in sign language discourse. *Journal of Pragmatics*, 30(5), 657–697.
- Lozano, S., Martin Hard, B., & Tversky, B. (2007). Putting action in perspective. *Cognition*, 103, 480–490.
- Marentette, P., Tuck, N., & Nicoladis, E. (2009). *Asynchronous gesture in American sign language narratives*.
- Markman, A., & Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Sciences*, 4(12), 470–471.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago, IL: University of Chicago Press.
- Parrill, F. (2009a). Dual viewpoint gestures. *Gesture* 9(3), 271–289.
- Parrill, F. (2009b). Interactions between discourse status and viewpoint in co-speech gesture. *Manuscript submitted for publication*.
- Parrill, F. (2009c). Linguistic and gestural viewpoint: Does a first person narrative evoke more character viewpoint? *Manuscript submitted for publication*.
- Parrill, F. (in press). The hands are part of the package: Gesture, common ground, and information packaging. In S. Rice & J. Newman (Eds.), *Empirical and experimental methods in cognitive functional research*. Stanford, CA: CSLI Publications.
- Parrill, F., Bullen, J., & Hoburg, H. (2009). Effects of input modality on speech-gesture integration. *Manuscript submitted for publication*.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.
- Quinto-Pozos, D. (2007). Can constructed action be considered obligatory? *Lingua*, 117, 1285–1314.
- Quinto-Pozos, D., Cormier, K., & Holzrichter, A. (2006). The obligatory nature of constructed action across 3 sign languages. *Workshop on Cross-linguistic Sign Language Research*. Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.
- Quinto-Pozos, D., & Parrill, F. (2008). Enactment as a communicative strategy: A comparison between English co-speech gesture and American Sign Language. *Workshop: Comparison of Signed and Spoken Languages*. Bamberg, Germany.
- Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, 4(5), 546–550.
- Stein, N. L., & Glen, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing*. Norwood, NJ: Ablex.
- Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, 35(3), 371–391.
- Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, 28(16), 81–88.
- Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 136, 168–171.

APPENDIX A

TABLE A1  
Criteria for coding discourse centrality (from Stein & Glen, 1979).

<i>Code</i>	<i>Type</i>	<i>Description</i>
P	Setting	Introduction of main characters
P	Internal response	The protagonist's reactions to the initiating event
P	Reaction	A response by the protagonist to the consequence
C	Initiating event	An action/happening that sets up a problem or dilemma for the story
C	Attempt	An action/plan of the protagonist to solve the problem
C	Consequence	The result of the protagonist's actions

TABLE A2  
Stimulus events and discourse status (C=central, P=peripheral).

<i>Event #</i>	<i>Description</i>	<i>Centrality</i>
R1	Rabbit winds up	C
R2	Rabbit throws ball	C
R3	Rabbit looks alarmed	P
R4	Ball flies out of stadium	C
R5	Rabbit chases out of stadium	C
R6	Rabbit takes bus to building	C
R7	Rabbit reads paper on bus	P
R8	Ball flies overhead	P
R9	Bus arrives at building, rabbit gets off	C
R10	Rabbit takes elevator to top of building	C
R11	Rabbit gets off elevator, runs across roof	C
R12	Ball flies by again	P
R13	Rabbit hooks self onto flagpole	C
R14	Rabbit pulls self up	C
R15	Ball flies by again...	P
R16	Rabbit throws mitt up	C
R17	Mitt catches ball	C
R18	Mitt falls back onto hand	C
S1	Skunk holds cat and kisses	C
S2	Cat escapes	C
S3	Skunk shrugs	P
S4	Skunk hops after	C
S5	Cat runs, banking turn on wall	C
S6	Skunk follows, hopping	C
S7	Cat goes up stairs, skunk follows	C
S8	Cat goes into room	C
S9	Cat closes door	C
S10	Cat leans against door, sighs with relief	P
K1	Dog walks up to house	C
K2	Woman complains about dog (taps foot)	P
K3	Dog comes in	C
K4	Kitten is on dog's back	P
K5	Dog puts kitten on shelf	C
K6	Dog covers kitten with bowl	C
K7	Dog agrees w/woman (nods)	P
K8	Kitten is going into room and dog follows	C
K9	Kitten is balancing on ball	C
K10	Kitten is playing with ball (tumbling on and off)	P
K11	Dog is laughing	P
K12	Kitten scrambles on ball, ball starts to roll	C
K13	Dog is alarmed	P
K14	Ball (with kitten on top) rolls towards woman	C
K15	Ball (with kitten) runs into woman's leg	P