UC San Diego

UC San Diego Previously Published Works

Title

Separating viewpoint from mode of representation in iconic co-speech gestures: insights from Danish narratives

Permalink

<https://escholarship.org/uc/item/5097j54n>

Author

FREDERIKSEN, ANNE THERESE

Publication Date

2017-01-10

DOI

10.1017/langcog.2016.35

Peer reviewed

**Separating Viewpoint from Mode of Representation in Iconic Co-speech Gestures:
Insights from Danish Narratives**

Anne Therese Frederiksen

UCSD Linguistics Department

9500 Gilman Drive

La Jolla, CA 92093-0108

Email: a.t.frederiksen@gmail.com

Abstract: During narrative retelling, speakers shift between different viewpoints to reflect how they conceptualize the events that unfolded. These viewpoints can be indicated through gestural means as well as through verbal ones. Studies of co-speech gestures have inferred viewpoint from gesture form, i.e. how entities are mapped onto the (primarily manual) articulators, but the merits of this approach have not been discussed. The present study argues that viewpoint is more than gestural form. Despite connections between the two, many other factors may influence a gesture's form. Assessing viewpoint from gesture form alone limits the applicability of gestural viewpoint as a window onto speakers' event conceptualization and introduces unnecessary differences in the categorization of viewpoint across gestures types. The present study examines iconic co-speech gestures in Danish narratives, and makes explicit the means used to infer gestural viewpoint. The approach advocated here ensures that the notion of viewpoint can be applied in a principled way to all or most iconic gestures.

Keywords: viewpoint, gesture, sign language, role-shift, mode of representation

Introduction

A fundamental part of the functioning of human cognition is that our perception is grounded in our bodies. This means that we perceive events and states in relation to our bodily experience and how our bodies are situated (e.g. Gibbs, 2006). Naturally, this tendency gets expressed when we communicate about the world. From the perspective of embodied cognition, perceiving and producing language involves mentally simulating the communicated content. For example, studies have claimed that humans understand language about actions and objects by mentally simulating those actions and objects (Zwaan, 1999; Kaschak & Glenberg, 2000; Stanfield and Zwaan 2001; Glenberg and Kaschak 2002; Zwaan, Stanfield, & Yaxley, 2002; Zwaan & Taylor, 2006; Masson, Bub and Warren 2008).

Scholars have described the perspective from which a speaker describes a given event with the term **VIEWPOINT** (e.g. DeLancey, 1981). Because viewpoint is a ubiquitous part of human cognition and communication, it plays a central role in the study of interaction and discourse (Borghi, Glenberg and Kaschak 2004; Brunye et al, 2009; Sweetser, 2012). By looking at a speaker's¹ viewpoint in an event description, we may glean insights into how he or she conceptualizes a given event.

As past research has revealed, speakers not only express their thoughts through verbal communication, but also through their gestural expressions (e.g. McNeill, 1992; Kendon, 2004). Accordingly, the speaker's viewpoint on an event also surfaces in the gestures that go along with speech (Parrill, 2012). By varying articulators and movement, speakers may change the gestural viewpoint of their utterances, although as with most gestures, the speaker may not be conscious of this type of variation. Studies of iconic co-speech gestures have used the way speakers represent referents to determine gestural viewpoint, but there has been little or no discussion of

the merits of this approach. As gesture has often been described as a window onto thought, it seems natural to assume that gestural viewpoint can inform us of aspects of the speaker's event conceptualization.

One model of gesture production, the Gesture as Simulated Action (or GSA) framework (Hostetter & Alibali, 2008, 2010), proposes that gestures arise from simulated action, and that two types of imagery, motor and visual, are responsible for gestures with different viewpoints. However, in a model of gesture production where viewpoint is a central feature, how we define viewpoint becomes a question of great importance. As viewpoint has traditionally been inferred based on gesture form, i.e. how entities from a narrative are mapped onto the manual articulators, this results in the assumption that different gestural forms are indicative of different kinds of conceptualization. In this paper, I argue that there is no one-to-one mapping between gesture form and viewpoint, because gesture form may depend on various factors such as event transitivity and general affordances. Consequently, although imagery may well play a role for gesture form, the assumption that gestural viewpoint is equivalent to gesture form means that viewpoint does not necessarily reflect event conceptualization. Viewpoint can and should be recognized in iconic co-speech gestures. However, in order to link viewpoint to cognitive processes such as event conceptualization, we need a clear definition of what it means for a gesture to have a viewpoint, and of whether and how viewpoint is connected to gesture form.

To achieve these aims, this paper proposes to separate the form properties of gestures from the way these properties can be interpreted. I provide a qualitative analysis of viewpoint examples from a dataset of elicited Danish narratives, exemplifying the various ways that gesture form and viewpoint interact. The approach presented in this paper has several advantages. First, it ensures that the notion of gestural viewpoint is relevant for understanding the speaker's event

Separating viewpoint from mode of representation in gesture

conceptualization and second, it allows for a more nuanced understanding of viewpoint that can be applied in a uniform manner to most or all iconic gestures. Finally, this approach to viewpoint makes testable predictions about cognitive processes that are different from predictions of current models.

In the following sections, I begin by discussing previous approaches to viewpoint in gesture and where existing categorizations might not suffice. I show that analyzing gestures according to their MODE OF REPRESENTATION (following Müller, 1998) provides a way of categorizing smaller parts of gesture ensembles' forms without invoking viewpoint (as also alluded to by Müller, 2014). On the basis of examples I then demonstrate an approach that allows us to interpret gestural viewpoints which do not necessarily derive from gesture form, and which do not rely on the animacy of the entities depicted in the gesture. This approach attributes special or privileged status to the speaker's body, as proposed by Engberg-Pedersen (2015), by assigning greater weight to the speaker's body than to the reference of other gestural articulators. Finally, the last section of the paper discusses the contributions and limitations of the study.

Viewpoint in gesture

The present study investigates iconic co-speech gestures (Kendon, 2004). Although both iconic co-speech gestures and deictic gestures can be viewpointed, my focus is restricted to iconic gestures. Such gestures represent content in different ways. Consider for example Figure 1a and 1b below. In Figure 1a, the speaker shows how a character in her narrative acts by using her own body as if it were the character's body. Conversely, in Figure 1b the speaker uses only her right hand to show the actions of a character, that is, she uses her hand to show the entity as a whole. In the gesture literature, this difference has generally been referred to as a difference in

Separating viewpoint from mode of representation in gesture

viewpoint (Beattie and Shovelton, 2001, 2002; McNeill, 1992, 2005; Parrill, 2009, 2010; Stec, 2012; Debreslioska et al., 2013) - most often framed as CHARACTER VIEWPOINT versus OBSERVER VIEWPOINT, although studies sometimes use different labels for comparable distinctions.²

Although the term viewpoint is often used in the co-speech gesture literature, exactly what constitutes a viewpoint is rarely specified. In studies that are concerned specifically with viewpoint (McNeill, 1992; Parrill, 2009), the definitions indicate that viewpoint is connected to the speaker's perspective on an event. The same studies' exemplification of viewpoint, however, suggests that the term refers to how the speaker uses her hands to represent referents³: when a speaker uses her own hands as a referent's hands, as in the cat-climbs-mountain gesture (Figure 1a), then the gesture is described as having a character viewpoint; when a speaker uses just her hand as an entire referent, as in the man-comes-down-stairs gesture (Figure 1b), then the gesture is labeled as having an observer viewpoint.

Figure 1a. Cat-climbs-mountain. Iconic gesture representing a cat climbing up a mountain.

Figure 1b. Man-walking-down-stairs. Beginning (left) and end (right) of an iconic gesture representing a man walking down stairs.

Separating viewpoint from mode of representation in gesture

Although the term viewpoint here would seem to refer to how the speaker uses her hands in the gesture, rather than to a cognitive phenomenon, the underlying assumption seems to have been that viewpoint can be inferred from gesture form. This is likely because of the intuitive appeal of a link between form and function. Clearly, we expect a connection between using one's body as if it were a character's and taking the viewpoint of that character; consequently, we might more generally assume some correlation between how the speaker uses her hands and the viewpoint she adopts on the event she is describing. However as mentioned, there has not been much discussion of the validity of such assumptions. In this paper, I argue that the choice of how to represent entities in co-speech gesture does not necessarily reflect the speaker's viewpoint or event conceptualization.

In the following sections, I discuss the paper's analytical approach including notions from sign language research that may inform and augment existing methods in gesture research, give details about the dataset and present examples that highlight key differences in viewpoint and mode of representation.

Describing gestural elements

Many studies have looked at the phenomena of viewpoint and perspective in sign languages across the world (e.g. Aarons and Morgan, 2003; Engberg-Pedersen, 1993, 1995; Liddell, 1998, 2003; Morgan and Woll, 2003; Perniss, 2007), and since both sign and gesture use the visual-manual modality, it is fruitful to turn to sign language research for additional analytical tools and terminology. Because gestures are spontaneous, non-conventional, and often holistic, difficulties arise, first with respect to pinning down how various aspects of a gesture contribute to its meaning, and second with respect to describing gestures in a consistent way. Unlike signing in a

Separating viewpoint from mode of representation in gesture

sign language, where the signs themselves provide the discourse context, a gesture's meaning generally depends on speech context, so much so that we are often unable to interpret a gesture until we have heard the accompanying speech. Various gestural aspects, however, can be determined formally. In the following sections, I first examine these formal properties and explain how gestural reference is determined. I then discuss how manual gestures can be classified as different modes of representation. Lastly, I define what constitutes a viewpoint in the present paper and discuss how viewpoint is assigned to gestures.

Formal properties of gesture, reference and mode of representation

The first analytical step in the present approach is to look at which articulators are involved in the gesture. An articulator (body, head, right forelimb, left forelimb and face) is considered involved in the gesture whenever it moves, or is maintained in a hold position, retaining a clear hand shape or position. I also include non-moving articulators, if the movement or positioning of other articulators relative to the articulator in question indicates that the speaker is treating the non-moving articulator as meaningful for the gesture. This occurs for example when another articulator moves towards the non-moving articulator (e.g. an entity mapped onto the speaker's hand moving towards the speaker's head or body), or when the speaker directs their gaze to the non-moving articulator.

I then identify the referent of each articulator. This is done by considering the form of the gesture and the entities and actions mentioned in concurrent speech. I also rely on the content of the stimulus materials to help determine a gesture's reference.

Having determined which articulators are involved in a gesture, the next step is to determine how the articulators are used symbolically. A number of schemes have been proposed

Separating viewpoint from mode of representation in gesture

for categorizing symbolic properties of gestures, that is how gestures represent entities (Streeck, 2009; Clark, 2016; see Kendon 2004 for an overview of earlier schemes). Here I follow the scheme proposed by Müller, who refers to this phenomenon as *MODE OF REPRESENTATION*. Mode of representation is defined as the way referents map onto the communicator's body, but because of the properties of the head, body and face (explained in more detail below), mode of representation is generally only specified for the hands or forelimbs. Müller distinguishes four modes of representation, namely *IMITATING*, *DRAWING*, *MOULDING*, and *PORTRAYING* (1998, pp. 323).⁴ She labels gestures as moulding and drawing when they show the shapes or outlines of objects, respectively. Portraying indicates that the entity in question is mapped onto the speaker's hand or hands in its entirety. Imitating gestures are those that the speaker uses to pantomime an everyday activity, like swimming or opening a window. Consider again Figure 1b above, the man-walking-down-stairs example. In Müller's categorization, this gesture is a portraying gesture, because the speaker maps an entire referent onto her hand.

Figure 2. Hold-frying-pan. Iconic gesture representing a mouse holding a frying pan.

Separating viewpoint from mode of representation in gesture

On the other hand, the handshape in Figure 2 is an imitating gesture because the speaker maps the hand of the referent onto her own hand to depict how a character in the narration handles an object. Similarly, in Figure 1a above, the cat-climbs-mountain gesture also exemplifies an imitating gesture. In this example, rather than depicting how an agent handles an object, the forelimbs of the cat are mapped onto the speaker's forelimbs, and there is no manipulation of an object involved.

To determine the mode of representation in a gesture, we need to look at each manual articulator separately. For purposes of this paper, I reduce Müller's four modes of representation to two, HANDS REPRESENTING HANDS and HANDS REPRESENTING ENTITIES (see Brentari et al, 2012 and Brentari et al, 2014 for use of similar terminology). Following the approach taken by Hostetter and Alibabli (2008, 2010), I assume that there is a link between type of imagery and different types of gestures. The GSA model proposes that motor simulation is responsible for gestures with character viewpoint, while visual simulation is responsible for gestures with observer viewpoint. The present paper assumes that the link is between imagery and mode of representation, not viewpoint, such that hands representing hands arises from motor simulation, and hands representing entities arises from visual simulation.

Consequently, all instances of imitation were treated as hands representing hands, because they map the referent's hand(s) onto the hand(s) of the speaker. Moulding, drawing and portraying count as hands representing entities, because they depict entities or parts of entities, either by the forelimbs of the speaker themselves, or through their movements.⁵ Note though that in the gestures discussed here there were no examples of moulding or drawing.

Although we cannot exclude the possibility that body parts such as the head or the legs can be used to portray other entities, most non-manual articulators tend to imitate when they are

Separating viewpoint from mode of representation in gesture

involved in a gesture. For this reason, I only include the manual articulators in the determination of mode of representation in the examples in this paper.

Importantly, the presence of a given mode of representation in a gesture does not necessarily mean that both manual articulators share a mode of representation, nor does it preclude other body parts from having a different reference from a manual articulator. Different entities can be mapped onto parts of a speaker's body simultaneously, but in different ways. For example, one hand may be used to portray, representing one referent in its entirety, while the other might imitate, i.e. the hand represents the same or a different referent's hand. Similarly the hands may represent an entity, while the body represents another entity, or does not participate in the gestural depiction. This phenomenon has been described as BODY PARTITIONING (Dudis, 2004), and it is a well-known phenomenon in signed languages. While the phenomenon also occurs in gesture, to my knowledge it has not yet been explicitly described.

Determining viewpoint

As mentioned above, studies of iconic co-speech gesture have often used the mode of representation in a gesture to infer the speaker's viewpoint on the event. Under such an approach, if a speaker uses his hands to imitate the way a character opened a window, then this leads to the inference that the gesture has a character viewpoint. Work on signed languages has identified certain combinations of linguistic elements (specifically classifiers) and perspective (or viewpoint) as prototypically aligned, but has also recognized that these elements may be in non-prototypical alignment as well - meaning that although the use of one type of classifier tends to indicate a certain viewpoint, this is not always the case (Perniss, 2007). It is very likely that there is in fact a similar tendency in gesture for each mode of representation to align with a specific

Separating viewpoint from mode of representation in gesture

viewpoint. There has, however, been no general discussion of the validity of this assumption among gesture researchers, nor any empirical data to show how often such prototypical alignment occurs.

In the present paper, I follow Müller (2014) in assuming that a given mode of representation does not necessarily imply one viewpoint. Müller discusses the relationship between her modes of representation and McNeill's distinction between what he refers to as character and observer viewpoint. She emphasizes that although the two phenomena, mode of representation and viewpoint, are linked, they are not identical, meaning that imitating gestures do not always have character viewpoints, and so forth. She gives the example of a gesture representing a knife cutting bread. Although the mode of representation here is portraying (referred to in Müller (2014) as REPRESENTING), this does not automatically give the gesture an observer viewpoint (Müller, 2014: 1698). Thus, the approach advanced in this paper, fleshed out in the following paragraphs, is closely related to Müller's in this respect.

The speaker's locus, expressive elements and gaze

The present analysis determines viewpoint based on the reference of the speaker's locus. The term locus is borrowed from sign language linguistics, and it refers to a point or a direction in the sign space, or in this case, gesture space (Engberg-Pedersen, 1993; Sandler & Lillo-Martin, 2004). The speaker's locus is the point in gesture space inhabited by the speaker's upper body. This locus may be used for the speaker herself, or the reference of the locus may be shifted, such that it represents another referent from the discourse. The reference of the speaker's locus is inferred by body orientation, gaze, by how the hands move in relation to the body, and sometimes by facial expression. As mentioned above, the properties of the non-manual

Separating viewpoint from mode of representation in gesture

articulators make them unlikely to represent other entities by way of portraying. Thus, the challenge for these body parts is to not determine the mode of representation, but whether they reflect the speaker's actions and emotions, or those of a character from the narrative.

Sign language researchers have shown that facial expressions (apart from those conveying grammatical information) made by the signer should not always be taken to represent the signer's emotions (e.g. Loew, 1984; Engberg-Pedersen, 1993, 1995, 1999; Lillo-Martin, 1995, 2012). Instead, facial expressions in a narrative may indicate the feelings of one of the entities being talked about. Engberg-Pedersen (1993, 1995) terms this use of facial expressions SHIFTED ATTRIBUTION OF EXPRESSIVE ELEMENTS. Similarly, speakers' facial expressions do not necessarily indicate the speaker's emotions (Fridlund, 1997). Instead, facial displays that occur in dialogue are generally used to show something about a referent (Bavelas and Chovil, 1997; Bavelas, Gerwing and Healing, 2014). In the examples below where I discuss gesture viewpoint, when facial expression is present, it helps identify which referent, if any, is represented at the speaker's locus.

Figure 3. Iconic gesture with gaze directed at the addressee.

Separating viewpoint from mode of representation in gesture

Similarly, the gaze of a speaker engaged in gesturing may be used to show something about a referent, rather than simply being the speaker looking around. In conversation, gaze is frequently directed at the addressee, as in Figure 3 above, but it may be directed virtually anywhere, for example to a point in space associated with a referent. Gaze may also imitate the gaze of a referent, as in Figure 4a below, where the speaker's gaze is similar to the character's in Figure 4b. The gaze may change in the middle of a gesture, and sometimes such a change is part of the speaker's representation of a referent, that is, the way the speaker's gaze moves represents the way the character's gaze moved in the stimulus. In the sign language literature, breaking gaze with the addressee is often taken as an obligatory indicator that a signer is role-shifting or constructing the action of a referent (e.g. Loew, 1984; Padden, 1986), that is, what the gesture literature has called imitating the actions of a character. Engberg-Pedersen (2015) explicitly links gaze to viewpoint. For her, if the signer's gaze is directed towards the addressee, the viewpoint cannot be fully with the character whose actions the signer is constructing (2015: 413). Studies examining specifically the correlation between gaze and viewpoint in gesture are, although still limited, beginning to emerge, and a number of observations regarding gesture type and gaze have been made. For example, Streeck (1993) discusses how speakers can use gaze patterns as a way of alerting the listener to a gesture's relevance. Similarly, Earis and Cormier (2013) show that speakers can use gaze as part of their depictions of story characters, just like signers of British Sign Language can. Sidnell (2006) and Thompson and Suzuki (2014) suggest that gestural reenactments by speakers tend to be accompanied by a break in gaze away from the addressee(s). However, it remains an empirical question whether in co-speech gesture research we can apply the strict criteria for gaze interpretation that are used in sign language research in the context of viewpoint. It is clear that signers and speakers generally use gaze in quite different ways (e.g.

Separating viewpoint from mode of representation in gesture

Lieberman et al., 2014), and unlike signers, speakers also rely on the verbal modality to support their narrative and provide information about which referent they are representing.

Figure 4a. Iconic gesture where the gaze direction is congruent with that of the character in the narrative.

Figure 4b. Still picture from stimulus clip. The cat's gaze is directed at the bird.

The properties discussed here make the referent of gaze harder to determine than the referents of arms and hands. Consequently, few studies of co-speech gestures include gaze (but see for example Sidnell, 2006; Stuckenbrock, 2014; Sweetser & Stec, 2016). However, procedures for analyzing gaze in sign language discourse are well established (e.g. Baker and Padden, 1978; Bahan and Supalla, 1995; Metzger, 1998). Recognizing its important role in determining viewpoint, the present study includes analysis of gaze. However, less strict criteria are adopted than in the sign language literature, for when gaze signals the different types of viewpoint.

It is not only the additional information in facial expression and gaze that indicates whether a referent is represented (and which one) at the speaker's locus. The direction of movement of the articulators is another source of information about this. For example, for

Separating viewpoint from mode of representation in gesture

Danish Sign Language Engberg-Pedersen gives the example of a signer describing a sleeping man being approached by a police officer (2015: 417). In this example, the police officer is represented with an entity classifier moving towards the signer's body. The sleeping man is represented at the signer's locus, but this is not visible based on articulatory shape of the hands or based on facial expression. The only indicator that the man is represented at the signer's locus is the fact that the articulator representing the police officer clearly moves with reference to the signer's locus. The articulators' direction of movement is categorized by looking at the axis of movement, how this relates to the rest of the body, and, in narratives elicited from visual stimuli, also at how the direction of movement corresponds to an observed action (Perniss, 2007; Perniss & Özyürek, 2008; Brown, 2008).

In Danish Sign Language discourse, the signer's locus has been argued to be the crucial factor in determining viewpoint or perspective assignment (Engberg-Pedersen, 2015: 419). As explained below, this paper adopts a similar approach.

Viewpoint

As I have pointed out, previous studies' approach to gestural viewpoint has been to equate it with mode of representation. The present study argues that it is beneficial to recognize the limitations of inferring viewpoint from mode of representation only, without establishing by which mechanisms this link between kinesic expression and viewpoint is brought about. At the maximum, doing so reduces the notion of viewpoint from providing conceptual information to being a mere description of form. At the minimum, we are narrowing our scope of inquiry, because we become excluded from recognizing actual viewpoint in a number of less typical

Separating viewpoint from mode of representation in gesture

gesture cases, which are discussed in later sections of the paper. Consequently, in this paper I make explicit how I link kinesic gesture form with gestural viewpoint.

Specifically, I separate mode of representation from viewpoint by following Engberg-Pedersen (2015) in giving privileged status to the speaker's locus, as determined by the speaker's body and gaze. Although it is not always this clear cut, the general idea is that if a speaker represents a referent at their locus in a gesture, then the viewpoint is a character viewpoint. If there is no referent represented by the speaker's locus, then the gesture has an observer viewpoint. Although the reference of the speaker's locus is tightly linked to viewpoint, the two notions are not fully overlapping. The reference of the speaker's locus simply identifies who inhabits the speaker's locus, while viewpoint categorizes this reference depending on what it tells us about the position from which the scene being described.

To see the difference between determining viewpoint based on mode of representation of the moving articulators vs. the reference of the speaker's locus, consider again the police officer example discussed above. Although this example was signed, we can apply the same principles to gesture. In this example, there was only one moving articulator, which was a hand representing a whole entity, namely the police officer. In an approach that determines viewpoint from mode of representation, a gesture of this type is classified as an observer viewpoint gesture, because of the hand representing a whole entity. Under the present approach, however, despite the fact that only the police officer is being represented with a moving articulator, the viewpoint is not with the police officer, nor with the narrator. Because there is another referent represented at the speaker's locus, namely the sleeping man, this type of gesture is classified as a character viewpoint gesture.

Separating viewpoint from mode of representation in gesture

Determining viewpoint on the basis of the reference of the speaker's locus rather than mode of representation allows for greater flexibility in assigning viewpoint to cases where multiple modes of representation are involved. Note however, that in this analysis, it is not necessary for the speaker to break gaze with the addressee in order for a gesture to have character viewpoint. Instead, I use the criteria that the gaze has to be, minimally, ambiguous between being the speaker's gaze to the addressee, and representing a character.

To summarize, the analysis in the present paper distinguishes two modes of representation, hands representing hands and hands representing entities, and two types of viewpoint, character and observer viewpoint. Mode of representation is determined on the basis of the symbolic value of the manual articulators, whereas viewpoint depends on the reference of the speaker's locus, which is determined by the speaker's body, gaze, and the direction of movement of the articulators. I group gestural viewpoint into the two well-known categories character viewpoint and observer viewpoint. Essentially, the gestural viewpoint is a character viewpoint, if the reference of the speaker's locus is to a character from the narrative, rather than to the speaker himself. Conversely, when the speaker is not representing a referent at his locus, then the speaker-as-narrator is observing events in the story, and the gesture has an observer viewpoint. Table 1 provides an overview of these possibilities. Following Perniss (2007), I will talk about alignment between mode of representation and viewpoint as prototypical in the case of hands representing hands together with character viewpoint and hands representing entities together with observer viewpoint, and non-prototypical alignment otherwise, including cases where the two modes of representations co-occur in the same gesture, or where viewpoint is mixed.

Table 1. Overview of Labels for Narrator Behaviors

	Label	Definition	Behavior
Mode of Representation	Hands Representing Entities	Speaker's forelimbs do not represent a referent's forelimbs	Speaker depicts an entity from the narrative with one or more articulators
	Hands Representing Hands	One-to-one mapping between speaker's and a character's forelimbs	Speaker maps a forelimb of an entity from the narrative onto their own corresponding forelimb
	Observer Viewpoint	Speaker's locus is not inhabited by a referent	Articulators move without reference to the speaker's locus; gaze is generally to the addressee ⁶
	Character Viewpoint	Speaker's locus represents an entity from the narrative	Articulators move with reference to the speaker's locus; gaze is as character or ambiguous between character and narrator

Methods

The gestures examples discussed in the present study are drawn from a dataset of Danish narrative retellings. It contains approximately five hours of video recordings consisting of cartoon narrations and re-narrations by 24 native Danish speakers (3 male, 21 female). The participants were university students participating voluntarily. All had spoken only Danish during childhood, and all had taken foreign language classes in English and at least one other language (German, French and Spanish). Their ages ranged from 20 to 27 years. Participants came to their recording session in pairs. This procedure was adopted assuming that participants

Separating viewpoint from mode of representation in gesture

gesture more naturally in a comfortable, informal setting. (e.g. Parrill, 2010). In each pair, the participants were randomly assigned to narrator and interlocutor roles. The narrator watched and retold a series of silent film and cartoon clips as well as a picture story with five pictures⁷. The instructions were to tell their partner what happened with enough detail so that the partner could reconstruct and re-narrate each clip. The interlocutor listened to and then re-narrated each story to the first narrator in turn. For the purposes of this paper, I only draw on examples from the narrators' retellings.

The dataset contains a rich variety of gestures. However, in order to examine the possibilities of combining viewpoint and mode of representation, the discussion is confined to a small number of gestures that were chosen because they exemplify such possibilities. It should be noted that the data are limited to video and picture retellings, which might influence the variety that is present in the dataset.

Analysis

The following presents an analysis of six gesture examples. I begin by giving examples of the two types of prototypical alignment and then present examples to show how viewpoint and mode of representation may crosscut each other in non-prototypical alignment. Finally, I discuss gestures that vary in the animacy of the entities they depict, and the role this has played in previous work.

Prototypical Alignment

Hands representing hands + character viewpoint

Separating viewpoint from mode of representation in gesture

As discussed above, we might expect that the way a speaker maps referents onto her body should bear a relationship to the viewpoint she adopts on the event she is describing. I will begin by looking at an example with character viewpoint and hands representing hands, which can be considered a case of prototypical alignment between viewpoint and mode of representation. In Figure 5, the speaker's gesture depicts a cat drawing on a drawing board. This example is a typical example of the gesture type that previous work has called a character viewpoint gesture as well (e.g. McNeill, 1992). In this gesture, the speaker's hands, head, body and gaze are all involved in the gesture. Her hands imitate the cat's paws, and their movement in relation to the speaker's body shows that the hands are not partitioned off from the body in this gesture. The speaker's head, body and gaze and the way the speaker's hands move in relationship to the body all indicate that the cat is represented at the speaker's locus. Consequently, the gesture has a character viewpoint.

Figure 5. Cat-drawing. Iconic gesture representing a cat drawing figures.

Og den sidder og tegner
and it sits and draws
'and it [the cat] is sitting and drawing'

Separating viewpoint from mode of representation in gesture

Hands representing entities + observer viewpoint

The typical counterpart to the type of gesture discussed in the previous example is what has been called an observer viewpoint gesture. In the example shown in Figure 1b (repeated below with speech transcription as Figure 6), we see how there is prototypical alignment between hands representing entities and observer viewpoint. This gesture depicts a man walking down a flight of stairs. Only the speaker's right hand, portraying the entity as a whole, is involved in the gesture. The speaker's head and upper body are not involved in the gestural representation. This is because neither is moving, and because the movement of the speaker's hand does not suggest that the speaker has assigned head or body referential value despite their inactivity. Further, the speaker's gaze is not the gaze of a character. This means that her body is the narrator's body, and therefore her locus indicates only herself as the narrator. As such, this gesture has an observer viewpoint. Defining viewpoint in the gesture types illustrated by the two examples just discussed does not pose any particular challenge. This is because these gestures show a very clear alignment between their mode of representation and their viewpoint. It is easy, however, to find examples that are less clearly defined than the two gestures discussed above. The next examples I discuss show that it is in fact possible for mode of representation and viewpoint to crosscut each other.

Figure 6. Man-walking-down-stairs. Iconic gesture representing a man walking down stairs.

Separating viewpoint from mode of representation in gesture

Og går så også ned ad trap-en
and walks then also down along stair-DEF
'and then [the man] walks down the stairs, too'

Non-Prototypical Alignment

Hands representing hands + observer viewpoint

This example shows a speaker describing how a mouse and an elephant toss a pancake back and forth between them, using frying pans.

In Figure 7, only the speaker's right hand and arm are involved in the gesture. The speaker's right hand is held in a grasping gesture, imitating how the mouse and elephant grasp the frying pan, making this a case of hands representing hands. However, the right forelimb is partitioned off from the rest of the body - the speaker's trunk, head/face and gaze are not depicting a referent.

Figure 7. Pancake-ping pong. Iconic gesture representing a mouse and an elephant tossing a pancake.

og så ender det med at de står og spiller pingpong
and then ends it with that they stand and play ping pong

Separating viewpoint from mode of representation in gesture

til hinanden med pandekagen
to each other with pancake.DEF
'and then it ends with them standing and playing ping pong to one another with the
pancake'

Additionally, the movement of the hand is not congruent with the event as seen through the eyes of either the mouse or the elephant. Rather, the hand moves between two locations that we interpret as those of the mouse and the elephant. Finally, the speaker's gaze is clearly directed towards the addressee in a way that is not ambiguous between being a character's gaze and the speaker's own. Thus, although the speaker hands represent hands, her locus inhabits only herself. That makes this gesture one where hands representing hands co-occurs with observer viewpoint, which is one way that mode of representation and viewpoint can interact that does not constitute prototypical alignment.

Hands representing entities + character viewpoint

The second obvious possibility for crosscutting of mode of representation and viewpoint is for hands representing entities to occur with a character viewpoint. This possibility is not attested in the corpus investigated here, but it is easy to see what such a gesture would look like. The example discussed above of a police officer approaching a sleeping man is exactly such a depiction: the only active articulator is the entity classifier moving towards the signer's locus. Under the current analysis, this is a case of hands representing entities, but because the signer's locus is inhabited by the sleeping man, the viewpoint is a character viewpoint. Although this example came from a signed narrative, a similar gesture could easily have occurred in a spoken and gestured narrative as well.

Separating viewpoint from mode of representation in gesture

Hands representing hands + dual viewpoint

As a last example of non-prototypical alignment between viewpoint and mode of representation, I discuss a gesture type that has neither character nor observer viewpoint. Instead it has a dual viewpoint. Figure 8a shows a gesture occurring during a narration of a Charlie Chaplin movie clip. The speaker's narration describes how Charlie Chaplin has fallen into a harbor basin and another man standing on the docks is trying to pull Chaplin out of the water. The result, however, is that Chaplin instead manages to pull the other man into the water with him. The speaker's hands and arms are directed towards the left, and his fists are closed, as if holding something. Helped by our understanding of the cartoon scene and by the accompanying speech, we see that the speaker is using his hands to imitate the man's actions as he tries to pull Charlie Chaplin out of the water.

Figure 8a. Pull-out-of-water. Iconic gesture representing a man pulling another man out of the water.

Figure 8b. Still picture from stimulus clip. The man tries to pull Chaplin out of the water, but instead Chaplin pulls in the man.

Den	første	mand	hjælper	den	anden	mand	op	af	bassin-et
the	first	man	helps	the	other	man	up	of	basin-DEF
'the first man helps the other man up from the basin'									

Separating viewpoint from mode of representation in gesture

Thus, the speaker's hands represent the man's hands. What the speaker's body refers to, however, is less clear. On the one hand, the body moves as the hands move, towards the speaker's left, which suggests that the reference of the speaker's locus is the man. On the other hand, the body seems to represent the narrator, because its movement is limited. Had the body not moved, this would suggest that the hands were partitioned off from the body, and that the speaker's locus was not inhabited by the man. This is supported by the speaker's gaze, which is clearly directed to the addressee in a way that is not ambiguous between character and narrator gaze. Moreover, in the gesture as it occurs here, the direction of movement of body and hands is not congruent with the referent's actions in the stimulus, in which the man is moving his arms forwards on the sagittal axis (compare Figure 8b). In Figure 8a, however, the speaker's gesture moves leftward on the lateral axis in a way that is congruent with the spatial arrangement of the characters in the stimulus clip as seen by the viewer.

Following traditional methods of determining viewpoint (McNeill, 1992; Parrill, 2009) results in the pull-out-of-water gesture (Figure 8a) being characterized as a character viewpoint gesture. This is because these studies determine viewpoint based on whether or not the speaker maps a character's hands onto his own hands, which the speaker does in the pull-out-of-water example. I would argue, however, that this analysis is not entirely satisfactory, as it fails to capture the fact that the speaker is moving his hands on the lateral axis, unlike the character in the stimulus. Under the present analysis, what I call the viewpoint is not only with the character in this gesture. Although the speaker's body functions primarily as the body of the character, which implies a character viewpoint, there is a simultaneous overlay of an observer viewpoint on the scene. This is caused by the speaker's body and hands moving in a direction that is incongruent with the action from the character's point of view, but congruent with the action as it

Separating viewpoint from mode of representation in gesture

appears from an observer's point of view. Further, despite hands representing hands being the only mode of representation in the gesture, the speaker's gaze does not imitate the character's. Thus, in this example the reference of the speaker's locus is ambiguous between character and narrator. The fact that the speaker's body and hands move together suggests that the speaker's locus represents the man, but the direction of movement, along with the gaze to the addressee suggests that the speaker's locus is occupied by the speaker-as-narrator. Unlike the pancake-ping pong example above where the hands were clearly partitioned off from the speaker's body, the involvement of the speaker's body in the pull-out-of-water gesture makes the reference of the speaker's locus ambiguous and gives the gesture a dual viewpoint, rather than an observer viewpoint.

At least one alternative possibility exists. It is possible that the gesture simply has a character viewpoint, and that the incongruent direction is due to the speaker avoiding a gesture direction that could interfere with the interpersonal space between himself and the addressee. However, this scenario requires the alignment between the leftward direction in the stimulus and the gesture to be purely coincidental. Until we have empirical studies that investigate the potential interaction between viewpoint and interpersonal space, we cannot know if this alternative analysis is a likely possibility. Further, the speaker's gaze is clearly to the addressee and cannot be interpreted as congruent with the speaker's locus representing only the man under the present analysis. Therefore, I maintain the analysis of an ambiguous reference for the speaker's locus, and consequently classify this gesture not as a character viewpoint gesture, but as a dual viewpoint gesture.

Animacy and Viewpoint

Separating viewpoint from mode of representation in gesture

Besides mode of representation, the concept of animacy has also played a role in how viewpoint has been determined in previous work by limiting which gestures have been considered candidates for having certain viewpoints. Below, I discuss examples that illustrate why, under the present approach, the animacy of a referent does not impact viewpoint or mode of representation.

Hands representing hands & hands representing (inanimate) entities + character viewpoint

Consider the gesture in Figure 9a. Here, the speaker is narrating an event involving a mouse baking pancakes. The mouse attempts to flip a pancake in the air, but instead of landing in the pan, the pancake falls on the mouse's face. As in the cat-drawing example above (Figure 5), the speaker's head, body and gaze are all involved in the gesture.

Figure 9a. Flip-pancake. Iconic gesture representing a mouse flipping a pancake.

Figure 9b. Still picture from stimulus clip. The character's gaze is directed at the pancake.

Separating viewpoint from mode of representation in gesture

Figure 9c. Cat-falls-into-arms. Iconic gesture representing a cat falling into the arms of a skunk.

Figure 9d. Still picture from stimulus clip. The skunk catches the cat.

Speech co-occurring with Figure 9a:

Men så får han den lige i hoved-et
but then gets he it exactly in head-DEF
'but then he [the mouse] gets it [the pancake] right in face'

Speech co-occurring with Figure 9c:

Og så falder øh så falder katten selvfølgelig lige
and then falls uh then falls cat.def of course exactly

i dens arme
in its arms
'and then um then the cat of course falls right into its [the skunk's] arms'

The speaker's right arm represents the mouse's right arm, and the speaker's gaze (compare 9a with the gaze of the mouse in the cartoon event that the speaker is retelling, Figure 9b), and body movements lets us infer that the mouse is represented at the speaker's locus. But what about the speaker's right hand? The hand is open, and the speaker moves it palm-up towards her forehead. Based on her speech and a comparison with the stimulus event, we understand that the speaker's right hand portrays the flying pancake in the flipping event, just before it hits the mouse's face. This makes the pancake an additional entity, besides the mouse, in the representation. Whereas for the speaker's right arm, we have hands representing hands, for the right hand we have hands representing entities.

With two types of representation involved in one and the same gesture, how do we determine the gesture's viewpoint? An argument made in previous work is that the pancake is an inanimate entity. As such it cannot have a point of view, and the gesture therefore has a character viewpoint (Parrill, 2009). While the gesture has a character viewpoint under the present analysis

Separating viewpoint from mode of representation in gesture

as well, there are two reasons why animacy is not a satisfying explanation for why this is. One reason is that the same kind of gesture may occur with an animate instead of an inanimate entity represented by the portraying articulator, as we will see below. The other reason is that, under the view of the present study, gestures representing inanimate entities also have viewpoints. If no entity is represented at the speaker's locus, such gestures have observer viewpoints. However, if the speaker personifies the inanimate entity and represents it at her own locus, even a gesture about an inanimate entity can have a character viewpoint.

Hands representing hands + hands representing (animate) entities + character viewpoint

The example in Figure 9c is taken from a narrative sequence in which the speaker is explaining how an enamored skunk is standing with its arms open, ready to catch a cat that is flying towards it. Like Figure 9a, this gesture represents two entities, but where the speaker in Figure 9a depicted an animate and an inanimate entity, in Figure 9c the speaker depicts two animate referents, as shown in Figure 9d. The speaker's left hand and arm are imitating the skunk's limbs and its handling actions. The speaker has directed her gaze towards her left arm, possibly displaying an interpretation of the character's, the skunk's, gaze. This, and the movement of the speaker's trunk and head signify that the skunk is represented at the speaker's locus. In addition, however, the speaker's right hand and arm work together to portray an entity, the flying cat. This gestural ensemble depicts two animate characters simultaneously, as opposed to the flip-pancake example (Figure 9a) in which the speaker depicted one animate and one inanimate entity.

Despite this difference, under the present analysis the two gestures are quite similar, yet they have been treated differently in terms of viewpoint (Parrill, 2009). Under the analysis proposed here, viewpoint depends on the speaker's locus, not on mode of representation or the

Separating viewpoint from mode of representation in gesture

animacy of the represented entities. This allows for treating these two gestures the same way, and in fact, they both have a character viewpoint. This differs from Parrill's (2009) analysis of gestures like the one in the cat-falls-into-arms example (Figure 9c), which she classifies as having dual viewpoint. Under the present analysis, the speaker depicts a referent with hands representing hands, both in the case of the cat-falls-into-arms gesture (Figure 9c), and the flip-pancake gesture (Figure 9a). The speaker's gaze depicts the character's gaze in both cases, which supports the speaker's locus as referring to the character. This leads to an interpretation of a character viewpoint and not an observer or a dual viewpoint. Lastly, in both cases the portraying of an additional entity, whether animate or inanimate, functions as an accessory to the depiction of events as seen from the point of view of the character.

Summary

In this section, I have shown how viewpoint and mode of representation interact in six different gestures. Table 2 provides an overview of how these two constructs can crosscut each other, exemplified with gestures discussed in the previous section. They show that although mode of representation may be indicative of viewpoint, this is not always the case. Viewpoint and mode of representation can align prototypically as either hands representing hands together with character viewpoint (as in Figure 5, the cat-drawing example) or as hands representing entities together with observer viewpoint (as in Figure 6, the man-walking-down-stairs example).

An advantage of this present approach is that it allows us to go beyond the affordances of the event being described. As will be discussed in more detail below, some scenes encourage the use of one particular mode of representation, and they are difficult to represent visually without resorting to that representation. This does not, however, necessarily imply that the speaker's

Separating viewpoint from mode of representation in gesture

viewpoint on the scene is aligned in a prototypical way with the mode of representation, as we saw in both Figure 7 and Figure 8a. Figure 7, pancake-ping pong, showed how hands representing hands can occur in a gesture with observer viewpoint. Figure 8a, the pull-out-of-water example, where the only mode of representation the speaker used was hands representing hands, had a dual viewpoint, not a character viewpoint. Although not all possible examples of non-prototypical alignment were observed in the present data set, we should not assume that they cannot occur.

The gesture examples discussed above also emphasized that the animacy of the depicted entity does not affect viewpoint. Because of this, a gesture depicting an inanimate referent may have the same viewpoint as a gesture depicting an animate referent. Similarly, both modes of representation, hands representing hands and hands representing entities, may co-exist in the same gesture, while the viewpoint is with only one entity. This, too, can happen both when the entity represented by the hands is inanimate (flip-pancake, Figure 9a), and animate (cat-falls-into-arms, Figure 9c). Leaving animacy outside of viewpoint assignment is another advantage of the approach adopted in the present study, because it allows for treating gestures with different articulator types the same with respect to viewpoint. As we will see in the discussion, there is no a priori reason to think that gestures representing animate entities necessarily arise from different cognitive processes than gestures representing inanimate entities.

Table 2. Interaction Possibilities between Mode of Representation and Viewpoint.

Viewpoint	Mode of Representation		
	Hands Representing Hands	Hands Representing Entities	Mixed Modes of Representation
Observer	pancake-ping-pong (Fig. 7)	man-walking-down-stairs (Fig. 6)	not observed
Character	cat-drawing	officer-approaching-man	flip-pancake

Separating viewpoint from mode of representation in gesture

	(Fig. 5)	(see Engberg-Pedersen, 2015: 417)	(Fig. 9a) cat-falls-into-arms (Fig 9c)
Dual	pull-out-of-water (Fig. 8a)	not observed	not observed

Discussion

The previous sections of this paper discussed how viewpoint can be separated from mode of representation and exemplified this by the analysis of different gestures. In this section, I situate this distinction in the historical context of gesture research, and discuss the benefits of the approach as well as the limitations of the study.

Some Viewpoint History

The analysis proposed in this paper hinges on taking seriously the fact that viewpoint in gesture is a complex notion with interactions between different layers of gestural function. This idea is not new in the field of gesture research. As one of the first studies to mention the notion of viewpoint, Cassell and McNeill suggested a differentiation between what they called VOICE and PERSPECTIVE. Their notion of voice refers to who is narrating at a given moment (1991, p. 388), which can be inferred from the form aspects of the gesture that the speaker uses, and the way the gesture is used spatially. This is tied together with how the narrator maps the depicted entity onto his or her own body. Two primary voices are recognized - character and observer/narrator voice. When the gesture is an enactment of a character's actions, the researcher infers that the voice is that of a character. If instead the speaker depicts the entire character on their hand, the voice being used is the narrator's own. Thus, the notion of voice appears identical to what later studies have referred to as viewpoint.

Separating viewpoint from mode of representation in gesture

Following Cassell and McNeill's definition, perspective is not identified by how the speaker maps the referent onto their body. Instead, this notion only applies to gestures with an observer's or narrator's voice. The perspective is internal if the narrator's gestures are made as if the narrator-as-observer is inside the event space. An external perspective implies that the narrator's gestures are made as if looking at the event from outside the event space. Thus, this work makes a clear distinction between perspective and voice, where there is some interaction between the two notions, although not all combinations are possible. This proposed distinction, however, was not generally adopted in the work that followed. For the most part, the two notions seem to have blended into one, viewpoint, which appears to be identified on the basis of Cassell and McNeill's notion of voice, that is, how the narrator's body represents entities, but is assumed to have (some of) the cognitive properties associated with their notion of perspective.

Consequences of Viewpoint Analysis

In this study, I have suggested that a detailed analysis of gestural components that goes beyond the form of the manual articulators, as well as of how these components interact, is crucial to understanding viewpoint. I have maintained that it is fruitful to distinguish between the mode of representation, and the actual viewpoint of the gesture.

As we have seen, it is certainly possible for the mode of representation and the viewpoint of a given gesture to align prototypically as either hands representing entities together with observer viewpoint, or as hands representing hands together with character viewpoint. However, by keeping the notions apart, we gain the flexibility to identify viewpoint that differs from mode of representation, and we are not required to assign viewpoint differently as a function of referent animacy. Under the analysis proposed here, a gesture where the hands represent entities may have either character or observer viewpoint, just as a gesture where the hands represent hands

Separating viewpoint from mode of representation in gesture

may have either viewpoint as well.

How we tackle questions of animacy and viewpoint impacts our understanding of wider theoretical questions. If viewpoint is seen as a reflection of conceptual processes occurring as language is being produced, then it is crucial that our understanding of viewpoint go beyond the speaker's chosen mode of representation. The separation proposed in this paper between mode of representation, and actual viewpoint in gesture not only has methodological consequences, but also has implications for theories of language production. The growing interest in linking language and perception to embodiment has also influenced thinking about gesture. The GSA framework claims that motor simulation generally leads to gestures with character viewpoint, and that visual simulation generally leads to gestures with observer viewpoint. However, two important issues regarding viewpoint have not yet been addressed by the GSA model.

The first concerns the idea of a straightforward one-to-one mapping of simulation type to gesture type. This issue was first raised by Parrill (2010), who argues against the implicit claim of the GSA framework that the two types of simulation and the corresponding two types of gesture are entirely separated from each other. Parrill points out that we need to account for what makes a speaker produce a gesture that contains both (what she calls) character and observer viewpoint, and as such shows evidence of arising from both motor and visual simulation. For example, a gesture like the cat-falls-into-arms gesture (Figure 9c) discussed above presumably involves motor simulation, because the speaker's arm is held out like the character's arm. However, an additional referent is introduced by the speaker's other arm representing an entity. According to the GSA framework, this second entity depiction should arise from a different kind of simulation than the first entity, namely visual imagery.

Gestures that have the same mode of representation but differ in viewpoint pose another

Separating viewpoint from mode of representation in gesture

problem for the GSA model. For example, the pull-out-of-water gesture (Figure 8a) involves a single representational mode, hands representing hands, but has a dual viewpoint. There is a clear difference between this type of gesture, and the same gesture made from a character viewpoint that cannot be captured by a difference in imagery.

Related to this is the question of whether the animacy of the depicted entity interacts with simulation, and how. Hostetter and Alibali do not address this question, but I would argue that the type of simulation the speaker engages in is (largely) independent of the animacy of the depicted entity. For example, just as it is possible for someone to imitate a character from a narrative, a speaker also has the option to use their entire body to imitate something inanimate, for example a tree, e.g. using the trunk of the body as the trunk of the tree and the arms as the branches. Thus, although different modes of representation may result from different types of simulation, there is no a priori reason to think that there are different cognitive processes underlying gestures that differ in animacy only.

The second and perhaps most important question concerns what viewpoint reveals about cognitive processes according to the GSA. Hostetter and Alibali do not define how they determine viewpoint, but their description suggests that what they call viewpoint corresponds to what I call mode of representation. The present claim is that there is a crucial difference between assuming a link between imagery and mode of representation as opposed to between imagery and viewpoint. I have argued against the traditional assumption that mode of representation directly indicates viewpoint, because there are many factors that can influence mode of representation.

As noted by Parrill (2010) verb transitivity plays a role in how the speaker maps a referent onto his or her own body, as does discourse structure. Recent work on reference tracking

Separating viewpoint from mode of representation in gesture

has shown that speakers prefer hands representing hands when they are maintaining reference, but use hands representing entities for referent re-introductions (Debreslioska et al., 2013).

Moreover, it seems that not all events can be represented equally well, or even at all, with either mode of representation (this phenomenon is discussed for gesture by Parrill, 2010, and for sign language by Özyürek & Perniss, 2010). For example, when describing a referent grasping or holding an object, a speaker can easily accompany the verbal description with a gesture where the hands represent hands. However, actions such as grasping and holding do not lend themselves well to depictions where the hands represent entities, because they rely on the hand of an agent. A detailed gestural representation of the crucial aspects of such an event, then, in fact necessitates mapping the character's hand onto the gesturer's hand, and renders the use of the hands to represent entities all but impossible. However, this is a limitation of the modes of representation, not of the narrator's conceptualization. Thus, the narrator may well conceptualize such an event from an observer's rather than a character's viewpoint. Crucially, this does not hinge upon the mode of representation that he or she employs for the gesture. Teasing apart viewpoint and mode of representation allows us to reserve the notion of viewpoint to truly reflect the narrator's conceptualization. This is important if we want to maintain the idea of viewpoint in gesture as a window into the speaker's mind. Moreover, separating these notions is a necessary step if we intend to maintain gestural viewpoint as a phenomenon with cognitive relevance comparable to other types of viewpoint.

An approach that equates viewpoint with mode of representation makes very different predictions about underlying cognition than does the present analysis. The former predicts that gestures that share a mode of representation, such as the cat-drawing and the pancake-ping pong gestures arise from the same cognitive processes. The present paper takes viewpoint to indicate

Separating viewpoint from mode of representation in gesture

from where a scene is viewed by the speaker, which may be unrelated to how she shapes her gestural articulators. Under this analysis, we assume that when a speaker uses a character viewpoint gesture, she is conceptualizing the scene as if she were in it, and consequently the elements of the scene should be conceptualized as near or close-up. Conversely, a speaker who uses an observer viewpoint gesture is assumed to have conceptualized the scene and its elements as seen from some distance away. Crucially, gestures with the same viewpoint are taken to indicate the same kind of conceptualization, independent of mode of representation. If we assume that a similar process applies to comprehension as to production, then gestures with the same viewpoint should evoke in the comprehender the same position in a scene, no matter their mode of representation. These predictions might be tested in a reaction time experiment that measures participants' responses to scene elements presented as near or far, after seeing scene descriptions involving observer/character viewpoint gestures that differ in mode of representation.

The present study has focused on providing a way to analyze consistently different combinations of viewpoint and mode of representation. No claim is made that the combinations discussed in this paper are exhaustive. Nonetheless, the argument is that the approach adopted in this paper, that is keeping the notions of viewpoint and mode of representation separate, allows for flexibility in classifying iconic gestures that can encompass the gesture types that were left outside the discussion in this paper. A limitation of this study is that it does not contain any quantitative information regarding how often mode of representation and viewpoint are in fact dissociated. The present paper has focused on laying out the theoretical importance of separating the notions, but future studies should assess the likelihood of different kinds of alignment to help us draw conclusions about the probability of each viewpoint occurring together with either mode

of representation.

Conclusion

This paper has examined the notion of viewpoint in iconic co-speech gestures. The goal of the paper is to show the benefits offered by a clear definition of how gesture form is linked to viewpoint. This allows for an expansion of the current use of the notion viewpoint in co-speech gesture studies to cover a wider range of gestures. I argued that the most typical use of viewpoint in the field is as a description of the form of the manual articulators. While there is a link between gesture form, which I have called mode of representation, and viewpoint, prototypical alignment between the two notions does not always occur. Borrowing from sign language linguistics, the analysis in the present paper teased these notions apart by looking closely at gestural articulators, facial expression, and gaze, and by regarding the speaker's locus as having a special status with respect to actual viewpoint. Though the analysis of gesture examples from Danish narratives, I showed that viewpoint and mode of representation interact in a variety of ways. Keeping these notions separate allows us to determine viewpoint consistently in different gesture types, because viewpoint does not depend on animacy of the intended referents of the gesture. I further outlined implications of viewpoint identification and classification for theories of gesture production and argued that the approach outlined in this paper restores the notion of viewpoint in gesture to its intended cognitive status. This paper has examined only a few gestural examples from one genre and one language. Although further work is necessary to establish the quantitative and cross-linguistic validity of the argument, the approach presented here provides an opportunity to re-examine and to better understand the effect as well as the importance of viewpoint in gesture studies.

Separating viewpoint from mode of representation in gesture

References

- Aarons, D., & Morgan, R. (2003). Classifier Predicates and the Creation of Multiple Perspectives in South African Sign Language. *Sign Language Studies*, 3(2), 125–156.
- Bahan, B., & Supalla, S. (1995). Line segmentation and narrative structure: A study of eye gaze behavior in American Sign Language. In K. Emmorey & J. S. Reilly (eds.), *Language, gesture, and space* (pp. 171–194). Hillsdale, NJ: Lawrence Erlbaum.
- Baker, C., & Padden, C. (1978). Focusing on the Nonmanual Components of American Sign Language. In P..A. Siple (ed.), *Understanding language through sign language research* (pp. 27–57). New York: Academic Press.
- Bavelas, J., & Chovil, N (1997). Faces in dialogue. In J. Russell & J. Fernández-Dols (eds.), *The Psychology of Facial Expression* (pp. 334-46). Cambridge: Cambridge University Press.
- Bavelas, J., Gerwing, J., & Healing, S. (2014), Including facial gestures in gesture-speech ensembles. In M. Seyfeddinipur & M. Gullberg (eds.), *From Gesture in Conversation to Visible Action in Utterance* (pp. 15-34).Amsterdam: John Benjamins.
- Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture*, 1(2), 129–149.
- 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(2), 179–192.
- Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, 32(6), 863–873.
- Brentari, D., Coppola, M., Mazzoni, L., & Goldin-Meadow, S. (2012). When does a system become phonological? Handshape production in gesturers, signers, and homesigners.

Natural Language & Linguistic Theory, 30(1), 1–31. <https://doi.org/10.1007/s11049-011-9145-1>

Brentari, D., Renzo, A. D., Keane, J., & Volterra, V. (2015). Cognitive, Cultural, and Linguistic Sources of a Handshape Distinction Expressing Agentivity. *Topics in Cognitive Science*, 7(1), 95–123. <https://doi.org/10.1111/tops.12123>

Brown, A. (2008). Gesture viewpoint in Japanese and English: Cross-linguistic interactions between two languages in one speaker. *Gesture*, 8(2), 256–276. <https://doi.org/10.1075/gest.8.2.08bro>

Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, 20(1), 27–32.

Bryant, D., & 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.

Cassell, J., & McNeill, D. (1991). Gesture and the Poetics of Prose. *Poetics Today*, 12(3), 375.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, 123(3), 324–347.

Debreslioska, S., Özyürek, A., Gullberg, M., & Perniss, P. (2013). Gestural Viewpoint Signals Referent Accessibility. *Discourse Processes*, 50(7), 431–456. <http://doi.org/10.1080/0163853X.2013.824286>

DeLancey, S. (1981). An Interpretation of Split Ergativity and Related Patterns. *Language*, 57(3), 626–657.

Dudis, P. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15(2), 223–238.

Separating viewpoint from mode of representation in gesture

- Earis, H., & Cormier, K. (2013). Point of view in British Sign Language and spoken English narrative discourse: the example of “The Tortoise and the Hare”. *Language and Cognition*, 5, 313-343
- Emmorey, K., Tversky, B., & Taylor, H. (2000). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation*, 2(3), 157–180.
- Engberg-Pedersen, E. (1993). *Space in Danish sign language: the semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum.
- Engberg-Pedersen, E. (1995). Point of view expressed through shifters. In K. Emmorey & J. Reilly (eds.), *Language, Gesture and Space* (pp. 133–154). Hillsdale, NJ: Lawrence Erlbaum.
- Engberg-Pedersen, E. (1999). Eye gaze in Danish Sign Language monologues: Forms, functions, notation issues. Paper presented at the 3rd Intersign Workshop, Siena.
- Engberg-Pedersen, E. (2015). Perspective in signed discourse: the privileged status of the signer’s locus and gaze. *Open Linguistics*, 1(1).
- Fridlund, A. (1997). The new ethology of human facial expression. In J Russell & J. Fernández-Dols (eds.), *The Psychology of Facial Expression* (pp. 334-46). Cambridge: Cambridge University Press.
- Gibbs, R. W. J. (2006). *Embodiment and Cognitive Science* (1 edition). Cambridge ; New York: Cambridge University Press.
- Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558–565. <http://doi.org/10.3758/BF03196313>
- Hostetter, A., & Alibali, M. (2008). Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin and Review*, 15(3), 495-514.

Separating viewpoint from mode of representation in gesture

- Hostetter, A., & Alibali, M. (2010). Language, gesture, action! A test of the Gesture as Simulated Action framework. *Journal of Memory and Language*, 63, 245-257.
- 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. <http://doi.org/10.1006/jmla.2000.2705>
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Key (ed.), *The Relationship of verbal and nonverbal communication* (pp. 207–227). The Hague/New York: Mouton.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press.
- 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.
- Liddell, S. (1998). Grounded blends, gestures, and conceptual shifts. *Cognitive Linguistics*, 9(3), 283–314.
- Liddell, S. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge/New York: Cambridge University Press.
- Lieberman, A., Hatrak, M., & Mayberry, R. (2014). Learning to Look for Language: Development of Joint Attention in Young Deaf Children. *Language Learning and Development*, 10, 19-35.
- Lillo-Martin, D. (1995). The Point of View Predicate in American Sign Language. In K. Emmorey & J. Reilly (eds.), *Language, Gesture and Space* (pp. 155–170). Hillsdale, NJ: Lawrence Erlbaum.

- Lillo-Martin, D. (2012). Utterance Reports and Constructed Action in Sign and Spoken Languages. In R. Pfau, M. Steinbach & B. Woll (eds.), *Sign language an international handbook* (pp. 365–387). Berlin/Boston: De Gruyter Mouton.
- Loew, R. (1983). *Roles and References in ASL: A Developmental Perspective*. Unpublished doctoral dissertation, University of Minnesota.
- Masson, M. E. J., Bub, D. N., & Warren, C. M. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language*, 59(3), 256–265. <http://doi.org/10.1016/j.jml.2008.05.003>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, D. (2000). Introduction. In D. McNeill (ed.), *Language and Gesture: Window into Thought and Action* (pp. 1–10). Cambridge/New York: Cambridge University Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Metzger, M. (1998). Eye Gaze and Pronominal Reference in American Sign Language. In C. Lucas (ed.), *Pinky extension and eye gaze: language use in deaf communities* (pp. 170–181). Washington, D.C.: Gallaudet University Press.
- Morgan, G. & B. Woll. (2003). The Development of Reference Switching Encoded Through Body Classifiers in British Sign Language. In K. Emmorey (ed.), *Perspectives on classifier constructions in sign language* (pp. 297–310). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Müller, C. (1998). Iconicity and gesture. In: S. Santi (ed.), *Oralité et Gestualité: Communication Multimodale, Interaction*, 321–328. Montréal/ Paris: L'Harmattan.

Separating viewpoint from mode of representation in gesture

- Müller, C. (2009). Gesture and Language. In: K. Malmkjaer (ed.), *Routledge's Linguistics Encyclopedia*, 214–217. Abington/ New York: Routledge.
- Müller, C. (2014). Gestural Modes of Representation as techniques of depiction. In: C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill and J. Bressemer (eds.) (2014). *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction*. (*Handbooks of Linguistics and Communication Science* 38.2.). Berlin/ Boston: De Gruyter Mouton.
- Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (ed.), *Language and gesture: Window into Thought and Action* (pp. 186–198). Cambridge/New York: Cambridge University Press.
- Padden, C. 1986. Verbs and role shifting in American Sign Language. In C. Padden (ed.), *Proceedings of the Fourth National Symposium on Sign Language Research and Teaching*, *Sign*, 44-57. Silver Spring, MD: NAD.
- Parrill, F. (2009). Dual viewpoint gestures. *Gesture*, 9(3), 271–289.
- Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5), 650–668.
- Parrill, F. (2012). Interactions between discourse status and viewpoint in co-speech gesture. In B. Dancygier & E. Sweetser (eds.), *Viewpoint in language: a multimodal perspective* (pp. 97–112). Cambridge/New York: Cambridge University Press.
- Perniss, P. (2007). Achieving spatial coherence in German Sign Language narratives: The use of classifiers and perspective. *Lingua*, 117(7), 1315–1338.
- Perniss, P. M., & Özyürek, A. (2008). Representations of action, motion, and location in sign space: A comparison of German (DGS) and Turkish (TID) Sign Language narratives. In J.

- Quer (Ed.), *Signs of the Time: Selected Papers from TISLR 8* (pp. 353–378). Hamburg: Signum.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12(2), 153–156.
- Stec, K. (2012). Meaningful shifts: A review of viewpoint markers in co-speech gesture and sign language. *Gesture*, 12(3), 327–360.
- Streeck, J. (1993) Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs*, 60(4), 275-299.
- Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Supalla, T. (1982). *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. Unpublished doctoral dissertation, University of California, San Diego.
- Sweetser, E. (2012). Introduction: Viewpoint and perspective in language and gesture, from the ground up. In B. Dancygier & E. Sweetser (eds.), *Viewpoint in language: a multimodal perspective* (pp. 1–24). Cambridge/New York: Cambridge University Press.
- Sweetser, E., & Stec, K. (2016). Maintaining multiple viewpoints with gaze. In *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities* (Vol. 55, pp. 237–258). Berlin, Boston: De Gruyter Mouton.
- Thompson, S. A., & Suzuki, R. (2014). Reenactments in conversation: Gaze and reciprocity. *Discourse Studies*, 1461445614546259.
- Zwaan, R. A. (1999). Situation Models The Mental Leap Into Imagined Worlds. *Current Directions in Psychological Science*, 8(1), 15–18.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168–171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology. General*, 135(1), 1–11.

<http://doi.org/10.1037/0096-3445.135.1.1>

Zwitserlood, I. (2012). Classifiers. In R. Pfau, M. Steinbach & B. Woll (eds.), *Sign Language An International Handbook*. Berlin/Boston: De Gruyter.

¹ The term *speaker* is used throughout this paper as a convenient way of referring to a speaker-gesturer.

² For different terminologies, see for example Kita and Özyürek (2003), Emmorey, Tversky and Taylor (2000) and Bryant and Tversky (1999).

³ Previous studies have included body parts other than the hands in their determination of viewpoint. It is important to note, however, that because the criterion for categorizing a gestures as having an observer viewpoint is that a referent is represented as a whole on the forelimbs (e.g. McNeill, 1992, 2005; Parrill, 2009; Debreslioska, et al, 2013), the difference between character and observer viewpoint in these studies essentially boils down to what happens on the hands.

⁴ Note that Müller's later work reduces these four modes of representation to two, namely ACTING and REPRESENTING (Müller, 2014).

⁵ I have restricted my focus to portraying and imitating, but indications of direction, such as path gestures, could be incorporated in the analysis as well. Note that the reduction of mode of representation into hands representing hands vs. hands representing entities does not directly correspond to Müller's (2014) distinction between acting and representing. Whereas the present paper treats molding and drawing as instances of representing the object, and hence as hands representing entities, Müller categorizes molding and drawing as acting.

⁶ Although as pointed out by a reviewer, the body may be an origo for the direction of the gesture, and the speaker may follow the gesture's path with their eyes.

⁷ The two film clips were excerpts from the silent Charlie Chaplin movie *City Lights*. Of the eight cartoon segments, one was the German *Die Sendung mit der Maus*, two were from *Canary Row*, and five were from *Pepe le Pew* (two from *A Scent of the Matterhorn*, two from *Two Scents Worth*, and one from *Little Beau Pepe*). The picture story was from a booklet of drawings constructed for another project (see Engberg-Pedersen, 2015 for details).