

Irene Mittelberg*

Experiencing and construing spatial artifacts from within: Simulated artifact immersion as a multimodal viewpoint strategy

DOI 10.1515/cog-2016-0124

Received October 25, 2016; revised March 30, 2017; accepted April 15, 2017

Abstract: Viewpoint has been shown to be a powerful construal mechanism in multimodal spoken and signed discourse, as well as in various other modalities and genres. This paper investigates embodied viewpoint strategies that have been observed when speakers combine speech, gestures, postures, gaze, and simulated action to describe their interaction with spatial artifacts such as gallery buildings, virtual architectural models, and paintings. Simulated artifact immersion is introduced as a multimodal viewpoint strategy whereby speakers submerge into their mental representation of an artifact by perceiving and experiencing it from an internal vantage point. It is argued that this viewpoint strategy tends to be employed when there is no narrative structure for the speakers to fall back on. The paper's aim is twofold: (a) to show that when speakers talk about their own experiences with spatial artifacts, distinguishing between immersed and non-immersed experiential viewpoint strategies may be more fitting than distinguishing between character and observer viewpoint; and (b) to discuss how considering the interaction of iconic, indexical, and metonymic principles in gesture may elucidate viewpoint phenomena in general.

Keywords: viewpoint, gesture, indexicality, metonymy, immersion

1 Introduction

As a universal and flexible construal mechanism, viewpoint shapes expressions and structures in spoken and signed discourses, as well as in other modalities and media. Being indexically anchored in dynamically evolving discourse structures, situated usage events reflect perspectivization (Cienki 2015; Langacker 1987; Sweetser 2012). Given the multimodal nature of human perception, cognition,

*Corresponding author: Irene Mittelberg, Institute of English, American and Romantic Studies, RWTH Aachen University, Aachen, Germany, E-mail: mittelberg@humtec.rwth-aachen.de

and interaction, the viewpointed conditioning of face-to-face interaction is not only reflected linguistically by the verb tenses and deictic elements employed, such as pronouns and demonstratives, but also by the manual gestures, head movements, eye gaze, and body postures that speakers use (e.g., McNeill 1992; Sweetser 2013). Over the past two decades or so, gesture research covering a host of different languages has contributed towards a fuller understanding of the central role that physical perspective and conceptual viewpoint play in shaping multimodal meaning construction or what we might call *multimodal usage events* (for recent reviews on viewpoint see, e.g., Bergen 2012; Dancygier and Sweetser 2012; Dancygier et al. 2016; Parrill 2012; Stec 2012).

This paper investigates viewpoint strategies that have been observed when speakers combine speech, manual gestures, body postures, gaze, and simulated action to describe and seemingly interact with spatial artifacts such as gallery buildings, virtual architectural models, and paintings. In contrast to animated cartoon stories, which have been amply used as stimuli in gesture research (e.g., McNeill 2000; see also Duncan et al. 2007), the artifacts of interest here are 2D or 3D spaces which, while exhibiting an inherent spatial organization, do not provide a narrative structure or story line for speakers to adopt for their descriptions. That is, the multimodal data used here do not involve retellings of witnessed film scenes or conversations. Accordingly, it seems necessary to reconsider the concepts of *character viewpoint* and *observer viewpoint* (McNeill 1992) for the discourse contexts of interest here. Advocating an embodied and multimodal understanding of language, cognition, and interaction, *simulated artifact immersion* is introduced here as a (typically) multimodally instantiated viewpoint strategy adopted, in particular, when language users are talking about their own previous or imagined experiences of spatial artifacts.

The goal of this paper is at least twofold. I will discuss how viewpointed processes of imagination and immersion interact with, and may also call into question, some of the established viewpoint strategies, depending on both the nature of the artifact and the nature of the multimodal discourse concerned. Furthermore, I will highlight how manual gestures, as a dynamic bodily modality intertwined with other modalities involved in language use, partake in embodied spatial cognition, construal operations, and creative processes (e.g., Cienki 2013; Cienki and Mittelberg 2013; Mittelberg 2013a; Müller 1998; 2014; Müller and Ladewig 2013; Sweetser 2007). To illustrate these points, I will first draw on video data consisting of multimodal German narratives produced by architecture students who simulated moving through imagined spatial structures and landscapes that they had either already experienced, designed, or saw emerging in front of their mind's eye while communicating their ideas to colleagues in the early stages of the design process. The second data set consists

of videotaped multimodal descriptions of paintings by Paul Klee in American English. Recalling the images from memory, participants elaborated on scenarios presented in the paintings by spontaneously enacting the postures and actions of the figures depicted. Crucially, they systematically shifted their viewpoint (e.g., Stec 2012) from the outside to the inside of the images.

Moreover, this paper pursues a theoretical interest in that it brings into the discussion a set of embodied principles and structures that seem to interact with viewpoint strategies in various ways. Notably, these are simulation (e.g., Gibbs 2006; Hostetter and Alibali 2008), the cross-modal interaction of iconicity and indexicality (e.g., Jakobson 1971; Mittelberg 2006; 2013a; Peirce 1960), and metonymy in gesture (e.g., Mittelberg and Waugh 2014). My intention here is to single out some of the motivating forces that seem to drive the complex cognitive and physical processes involved in imagining, conceptualizing, describing, and also creating (virtual) artifacts in an experientially grounded, multimodal fashion (e.g., Hutchins 2005; Johnson 2007; Turner 2006). Overall, this paper aims to contribute theoretical and observational insights that may strengthen multimodal accounts of viewpoint meaning construction in connection with other embodied principles of construal and understanding that are central to cognitive linguistics.

2 Versatile viewpoint strategies: Embodying multiple perspectives on the same event

As a bodily medium of expression, gestures always emanate from the viewpointed body of the speaker and hence cannot be understood and analyzed in isolation from the body in the way that speech can. At the same time, body postures, manual gestures, head movements, eye gaze, and facial expressions may also be used very effectively to display the behavior as well as the mental and affective states of others. This perspective-taking ability is comparable to how speakers or writers narrate stories from a first-person or third-person perspective as well as integrate both direct and reported speech into their narrative (e.g., Sweetser 2012; 2013).

Bodily articulations have their own modality-specific affordances and ways of describing, quoting, enacting, and thus contributing observable, action-based facets to storytelling and other discourse genres (e.g., Clark 2016; Mittelberg 2014; Müller 1998; 2014; Streeck 2009). Since they are produced less consciously than speech or written discourse, gestures may provide valuable insights into less-monitored aspects of construal and spatial cognition during communication

(e.g., Cienki 2013; Sweetser 2007). This section serves to lay out some of the major viewpoint strategies discussed in the literature and to introduce the viewpoint phenomena that are of special interest to the present proposal.

2.1 Gesture and viewpointed meaning construction

Gesture scholars have approached the semiotic complexity and multimodal dynamics of viewpoint phenomena from different angles. A fundamental issue to consider is how basic human cognitive capacities and the semiotic systems that constitute spoken interaction feed into the construction and expression of viewpoint. These concerns align with some of the main premises of cognitive linguistics, namely, that human perception, conceptualization, and language use are inherently perspectivized (e.g., Croft and Cruse 2004; Verhagen 2007) and multimodal (e.g., Cienki 2013; Dancygier and Sweetser 2012). Parrill (2012), for instance, distinguishes between *conceptual viewpoint*, on the one hand, and *linguistic* and *gestural viewpoint*, on the other, before further categorizing viewpoint expressed in gesture. Both linguistic and gestural viewpoint may, according to Parrill (2012: 110), reflect conceptual viewpoint.

With regard to gestural viewpoint, the ground-laying observation of McNeill (1992) was that a real-world experience or a movie scene may be described from an internal perspective by adopting character viewpoint or from an external perspective by assuming observer viewpoint. In other words, we may become the person or cartoon character we are talking about, thus iconically mimicking his, her, or its posture, actions, or way of speaking; alternatively, taking a step back, we can portray, for instance, how we saw two people engaged in conversation facing each other by using our two hands with open palms oriented towards each other to represent the respective heads of the conversational partners. Speakers may further use two distinct articulators to display two perspectives on the same event through *dual viewpoint* (McNeill 1992; see Parrill 2009; for an overview of dual viewpoint in gesture and American Sign Language [ASL]). For example, while representing two dialogue partners with our hands as just described (observer viewpoint), we can simultaneously assume the role of one partner by imitating her surprised or sad facial expression (character viewpoint). Hence, besides exhibiting subjective aspects of the gesturer's own bodily and cognitive disposition, or what s/he profiles as particularly relevant in the ongoing discourse, communicative action may likewise embody the intersubjective experiences of others, including their *emotional viewpoint* (Sweetser 2013).

A growing body of gesture research has shown that expressing a perspective on a given event or witnessed conversation is not limited to these major

viewpoint strategies. On the contrary, multimodal performances may become considerably more complex when multiple viewpoints are represented by distributing them across different body parts (see, e.g., Dudis 2004 on *body partitioning* in ASL). What is truly remarkable is the cognitive and semiotic versatility with which speakers recruit their head, eyebrows, facial mimics, manual gestures, whole-body postures, as well as eye gaze as viewpoint markers (e.g., Debras 2013; Liddell 2003; for eye gaze, see Brône et al. (2017); and Sweetser and Stec 2016). Indeed, speakers of different spoken and signed languages have been observed to enact simultaneously mixed or rapidly shifting viewpoints on one and the same event (see Stec 2012 for a review). These converging insights prompted Deliema and Sweetser (2016) to argue in favor of resolving the dichotomy of character and observer viewpoint and to advocate a multidimensional account of perspective-taking.

Examining the noted multidimensionality and creativity of viewpoint construction across modalities, Sweetser (2013) further addresses spatial, temporal, and social facets of embodied viewpoint in language, gesture, and sign language. She reports on a viewpoint analysis of the retellings of autobiographical incidents in American English (see also Sweetser and Stec 2016). One study participant not only imitates how a town hall official was holding an imaginary document while talking to her in the depicted scene, but she also simultaneously mimics her own head gesture and verbal response to the clerk. The same narrator, while maintaining the clerk's arm position, then blends the depicted scene with "her own current part in the Real Space conversation with her real-world interlocutor" (Sweetser 2013: 248). This qualifies as an "impossible gesture" (Sweetser 2013: 246) because, for a moment, it seems as if the clerk (represented by the narrator) is talking to the real-world interlocutor. Hence, while the narrator has only one body, she recruits different body parts to concurrently personify several imagined people. This is also an instance of gestural embedding: "The retention of the story-character's viewpoint seems frequently to be a way of visibly holding the floor for return to the narrative, even while granting the temporary floor to an interruption by the story-listener" (Sweetser 2013: 248–249). As this example shows, when describing their own real-world experiences, narrators may assume the role of their past selves and those of other people. Narrators thus tend to partially, that is, metonymically, enact or display some aspects of their own linguistic and physical behavior as well as that of story characters or conversational partners from their own past (see Mittelberg and Waugh 2014 on internal and external metonymy in gesture).

This raises a central question: What circumstances prompt speakers to adopt a certain viewpoint, to shift viewpoint, or to opt for a combination of two or more perspectives? Exactly how factors such as cognitive load, information

management, event structure, linguistic structure, discourse genre, and idiosyncratic properties of the different modalities interact in driving the choice of viewpoint, as well as how information and pragmatic functions are distributed across the vocal and gestural modalities, are issues requiring further investigation (e.g., Emmorey et al. 2001; Parrill 2009; 2010; 2012; Stec 2012). Yet gesture researchers have identified some tendencies as to how the multimodal representation of certain kinds of elements and actions may pattern with particular viewpoint strategies. For example, in American English narrations of *Canary Row* cartoons, Parrill (2009) has revealed the decomposition of manner and trajectory (or path; Talmy 2000) in the portrayal of a sliding movement detailed in the corresponding cartoon description text: one of the study participants used her right hand to trace the movement trajectory while using her left hand to iconically represent the way in which the movement of the character's hand was helping it to maintain a balanced posture (Parrill 2009: 284). This reflects a general trend to use observer viewpoint to portray movement trajectories of objects and people, or spatial constellations in general, and to opt for character viewpoint to imitate the manner of movement or bodily actions, postures, or facial expressions of cartoon characters or other observed agents (Debras 2013; Parrill 2010; Stec 2012).

Exploring multimodal descriptions of pictorial and architectural artifacts, and of imagined scenes inside them, this paper focuses on how gestures may offer modality-specific insights into physical, spatial, and emotional aspects of viewpoint construction during imaginative and communicative processes. Since this focus brings to the fore the speakers' own, real or imagined, experience with spatial artifacts, another viewpoint strategy becomes particularly relevant here, namely, *experiential viewpoint* (Sweetser 2013), which will be discussed in detail in Sections 3 and 4.

2.2 Illustrating multiple viewpoint construction: Experiential immersion

For a first illustration of multimodal viewpoint construction, let us consider the following description of a routinized motion event taken from the first data set used in this study (adapted from Mittelberg et al. 2017). It was produced by an architecture student, a native German speaker talking about a childhood memory of how every morning on her way to kindergarten she would run down the seemingly endless, winding stairs in her house. The entire episode is narrated in the past tense and from the first-person perspective (SG and PL). Imagining herself inside the house where she grew up, she opens her verbal

description by hinting at the angle from which her portrayal begins: “*und wir waren dann ganz oben*” [and so we were on the very top]. She physically enacts this perspective-taking through her body posture, head orientation, gaze and gestures by pretending to be looking down, with a neutral facial expression, from the top flight of the staircase: first, she holds up her left downward-facing, flat hand alluding to the top flight (Figure 1); then, the index finger of the same hand represents herself spiraling down the staircase (Figure 2). While saying “*könnte ich immer die Treppe von oben runtersasen*” [I could always run down the stairs from the top], she draws a spiral-like gestural trace starting at



Figure 1: Indicating the vantage point of immersed perspective-taking. “*Und wir waren dann ganz oben*” [And so we were on the very top].



Figure 2: Eye gaze as viewpoint marker inside an imagined space while the index finger spirals downward. “*könnte ich immer die Treppen von oben runtersasen*” [I could always run down the stairs from the top].

eye level, going downward and around six times, and ending at waist level (see the winding motion-capture trace in Figure 3). She then uses both hands, with their open palms facing each other, to draw two vertical lines going upward from the imagined ground level (Figure 5), thus outlining the shape of what she conceptualizes as an enormous tower (represented by the two vertical traces in Figure 3).

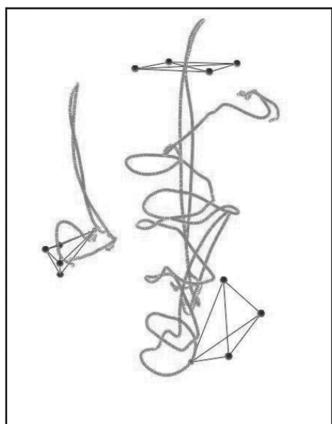


Figure 3: Motion-capture trace of gestures portraying a downward spiral motion event and the outer contours of a tower.

- (1) *Und wir waren dann ganz oben ... und immer wenn ich halt zum Kindergarten musste, dann konnte ich immer die Treppe von oben runtersausen, ganz schnell, aber diese Treppe nahm und nahm und nahm und nahm kein Ende, weil ich eben das Gefühl hatte, dass wir in einem riesengroßen Turm wohnten.*

'And so we were on the very top, and each time when I had to go to kindergarten, I would always run down the stairs from the top ... very fast, and the staircase didn't and didn't and didn't and didn't come to an end because I had the feeling that we were living in this enormous tower.'

What we see here could qualify as an instance of mixed viewpoint (Stec 2012) coming to bear via a partitioning of the body (Dudis 2004; see also Calbris 1990; on *body segments*): the speaker's body, head, and eye gaze show someone observing a scene from above; at the same time, her index finger portrays her own action from the observer viewpoint while she is verbally describing it in detail using the first-person singular. Upon closer scrutiny, however, it becomes clear that the concepts of observer, character, or dual viewpoint do not seem ideally suited to accounting for how this multimodal performance gets staged through the use of multiple viewpoints. Although the speaker is the person who

descended the staircase in the narrated event, she does not simulate the action of running down the stairs via character viewpoint, e.g., by leaning forward slightly with her torso, pretending to hold on to a handrail, or moving her arms or feet up and down; nor does she separate the path from the manner of the motion by using different articulators to create character and observer viewpoints, as described in Section 2.1 (Parrill 2009). Instead, the speaker's own experience of interacting perceptively and physically with the spatial configuration in question shapes how spatial and experiential viewpoints are enacted. The crucial point is that she displays the motion event's path in the form of a perspectivized spiral viewed from a location that is not situated outside the described scenario (as is the case for most cartoon-based path gestures) but from a vantage point that is indexically anchored inside the imagined, previously experienced spatial artifact (Mittelberg 2013a; Sweetser 2012).

The idea that the stairs seemed endless and the space appeared huge to the speaker when she was a child comes across effectively not only due to the spatio-temporal extension of the winding gesture and the concurrent iconic reduplication in her speech: "*die Treppe nahm und nahm und nahm und nahm kein Ende*" [the staircase didn't and didn't and didn't and didn't come to an end]. The portrayal is also enhanced, and staged, by how the speaker directs her gaze. As Sweetser and Stec (2016) have convincingly shown, gaze is a means to switch back and forth between the story space and the current narration space, or between the imagined space and real space (Dudis 2004; Janzen 2012; Liddell 2003; Parrill 2012; Sweetser 2007). This is also the case in our example, where the speaker intermittently looks at the interviewer. However, here, viewpoint creation through eye gaze particularly plays a constitutive role in mentally reconstructing the imagined spatial artifact and the motion event unfolding within it. This is afforded through cross-modally interacting iconic and indexical elements: first, the speaker directs her (indexical) gaze down onto her left hand performing the iconic spiral-like motion event (Figure 2). Here, her left index finger functions, via metonymy, as an *abstract body icon* (Mittelberg 2014) that is indexically pointing downward. Then, when talking about her feelings about the spacious volume, she seems to be looking all the way up to the top floor of the staircase, as if she were standing on the ground level of the building; she thus creates another index (Figure 4). In fact, these two spatially anchored viewpoints signaled by the speaker's gaze seem to condition the orientation of the speaker's head as well as the direction of her fingers and gestural movements, which are pointing or moving first downwards (Figures 1 and 2) and then upwards (Figures 4 and 5). It is when the speaker is drawing the tower's vertical contours iconically in the air (Figure 5) that, for the first time, she resorts to taking a frontal, external view on the house.



Figure 4: Eye gaze as the viewpoint marker inside an imagined space. “*weil ich eben das Gefühl hatte ...*” [because I just had the feeling ...].



Figure 5: Gaze directed at the interlocutor, hands tracing the building’s outer walls. “*dass wir in einem riesengroßen Turm wohnten*” [that we were living in this enormous tower].

Regarding linguistic indexical elements, the following deictic expressions, or *shifters* (Jakobson 1971), are instrumental in the above example of viewpoint construction: personal pronouns in the first-person singular and plural (*ich*, ‘I’; *wir*, ‘we’), spatial prepositions and adverbials (*ganz oben*, *von oben*, *runtersausen*, *in einem riesengroßen Turm*, ‘on the very top, from above, run down, in an enormous tower’). Moreover, the modal particles *hast* and *eben*, both

meaning ‘just’ or ‘simply’, function as pragmatic modifiers signaling a personal viewpoint or stance (Schoonjans 2014). For instance, through the use of *eben*, the speaker reinforces the idea that, to her, it felt as if her family were living in a gigantic tower: “*weil ich eben das Gefühl hatte, dass wir in einem riesengrossen Turm wohnten*” [because I just had the feeling that we were living in an enormous tower]. Smiling, the speaker accompanies this subjective statement with the bimanual palm-up open hand gesture shown in Figure 4. In this context, it may count as a predominantly pragmatic variant of what Parrill (2012: 102–103) calls a *metanarrative gesture* expressing *narrator viewpoint*; here, it occurs in the moment the speaker reestablishes eye contact with her interviewer. Her amused, joyful facial expression reflects her own emotional viewpoint on this positive childhood memory.

Based on these first observations, we now are in a position to characterize multimodal instances of experience-based viewpoint construction in a more principled way.

3 Simulated artifact immersion as a viewpoint strategy in multimodal spoken discourse

As the literature on viewpoint in language and gesture reviewed in the previous sections shows, multimodally instantiated viewpoint strategies are commonly observed, regardless of whether the depicted scene is something that the speakers participated in, belongs to a story told to them, or is part of an artifact such as a film, a cartoon, or a novel. One of the central questions is how the nature of a given artifact, scene, or experience influences what kinds of viewpoint strategies can be expected to shape the respective discourse and also the kinds of gestures employed. As already emphasized, the present proposal hopes to contribute insights to this discussion by exploring multimodal descriptions of the imagined experience of both real-world and virtual spatial artifacts. Simulated artifact immersion will be defined and further exemplified in this section.

3.1 Defining simulated artifact immersion

Simulated artifact immersion is understood here as a viewpoint strategy by which speakers submerge themselves in their mental representation of an artifact by perceiving, experiencing, and conceptualizing it from an imagined internal vantage point for communicative purposes. In this paper, our explorations are limited

to spatial artifacts such as 2D painted scenes, virtual architectural models, and physically existing 3D buildings. Compared to most of the research on viewpoint in co-speech gesture discussed above, the decisive difference here is that the speakers do not have an existing narrative structure to fall back on, for example, in the form of a story line with a plot and characters. Specifically, they do not need to stick to a given script or some other kind of given temporal progression of events; they are free to selectively describe as well as imaginatively elaborate certain aspects of the scenarios offered by an architectural spatial structure or a painting that they are talking about.

A decisive point to draw from the discussion of example (1) is that the concepts of character, observer, and dual viewpoint do not suffice to account for how the speaker describes a personal experience by immersing and positioning herself in specific locations inside the imagined space from where she performs differently viewpointed, physically displayed acts of perspective-taking (cf. Section 2.2). The very fact that she only briefly assumes a frontal, neutral perspective reveals her subjective experience-based view of (and feelings about) the particular space and motion event she is describing. With her gestures and eye gaze, she construes the dimensions of the spatial artifact in question and represents the motion event that happened repeatedly within it, each time scaled down to the dimensions of her gestural articulators and her gesture space. We can further conclude from these first observations that the indexicality inherent in a speaker's gaze, deictic gestures, as well as in her/his iconic postures and gestures may provide observable indices that point back toward the vantage point from where s/he is conceptualizing and describing an imagined space, event, or any other kind of experience. What is crucial here is the insight that "imagined perception is viewpointed" (Sweetser 2013: 239) as well. The following embodied phenomena are particularly relevant regarding multimodal instantiations of simulated artifact immersion and will be briefly discussed below: immersion, experience, imagination, and simulation.

Immersion is a general cognitive mechanism that can be triggered by linguistic and non-linguistic space builders (e.g., Coulson 2001; Ehmer 2011; Fauconnier and Turner 2002). When retelling a film or an animated cartoon story, immersive processes are involved as soon as a narrator puts herself in the position of a character and describes an event from the inside, for instance, by using character viewpoint (e.g., Bergen 2012). As described in Section 2, this also applies, to varying degrees, to cases where a speaker is recounting, and thus re-living, a dialogic exchange in which she herself took part. When looking at how such immersive techniques play out in written narratives, we see comparable effects, especially regarding the use of verbal tenses and aspect (see, e.g., Dancygier 2012; Vandelanotte 2009; on the progressive in English). For example, when reading a

passage in a French novel, we would experience the narrated situation differently depending on whether it contained predominantly imperfective or primarily perfective verb forms (see also Matlock et al. 2012). As Sweetser points out,

imperfective verb forms evoke *immersion* in (*internal experience of*) an ongoing situation, while perfective forms evoke external viewpoint on a situation as a whole (...). The impression given by the imperfective past is one where the narrator's viewpoint is still present (...) but is 'immersed' in the local *imagined experiential viewpoint* of the character. (Sweetser 2013: 244) [emphasis mine]

Imagined experiential viewpoint (Sweetser 2013: 244) proves to be of special relevance for the concept of simulated artifact immersion in the context of multimodal interaction, as it highlights temporal and dynamic aspects of experience and imagination in the light of semiotically mediated scenarios.¹

Besides spatial and temporal dimensions of imagined experience, connected aspects of mental simulation and sensorimotor programs are of central importance here (e.g., Bergen 2012; Gibbs 2006; Gibbs and Matlock 2008), for speakers often simulate actions with their bodies in dynamic multimodal descriptions (e.g., Cienki 2013; Müller and Ladewig 2013; Perlman and Gibbs 2013). Previous gesture research has addressed the question of whether scenarios personally experienced by the speaker prompt the use of character viewpoint. For example, within Hostetter and Alibali's (2008) gesture-as-simulated-action (GSA) framework it is maintained that this is usually the case, whereas Parrill (2010) argues that people's strong propensity to simulate actions inclines speakers to act as if they have experienced a situation even if they have not (see Marghetis and Bergen 2014 for a recent review on gesture and mental simulation). In what follows, this tendency to simulate one's own actions and the actions of others will be further explored in light of immersive processes that are evoked when speakers are talking about their experience of both physically inhabitable and uninhabitable spaces. First, the difference between non-immersed and immersed instances of multimodal viewpoint construction will be exemplified.

3.2 Non-immersed viewpoint construction from the outside of spatial artifacts

Let us first consider an example of non-immersed viewpoint construction manifested in a multimodal description of a spatial artifact. The gestural performance in Figures 6a-c iconically portrays aspects of a specific building

¹ For research on online processes of viewpoint construction in mainly written text, see also Dancygier (2012), and MacWhinney (2005).

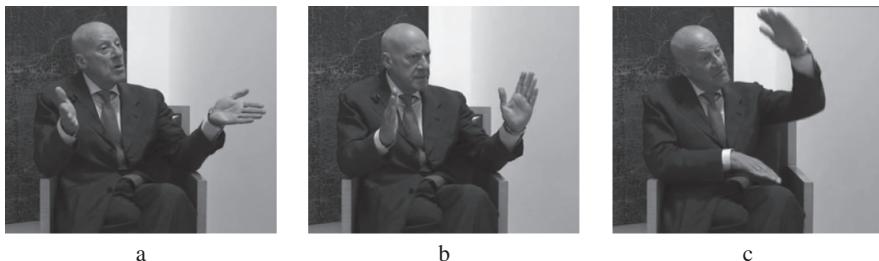


Figure 6: Non-immersed viewpoint construction from the outside of a spatial artifact.
“(a) *It's a twenty-five foot wide slot ...* (b) *so it's very tight ...* (c) *that means that it's a vertical gallery*”.

that the speaker has not only experienced many times but also conceptualized and designed (taken from an interview on ArchRecord TV, January 5, 2011). In (2), the British architect Norman Foster is describing the Sperone Westwater Gallery in Manhattan, pointing out its characteristic feature, namely, its comparatively narrow and high gestalt. To describe its spatial dimensions, Foster brings this enormous building down to human scale by using his hands to demarcate three differently oriented chunks of space: first, on “*it's a twenty-five foot wide slot*” his hands, with the open almost vertical palms facing each other, are held a little more than shoulder-width apart. Indicating the width of the building slot, the hands open slightly upward, and the fingertips turn slightly outward and toward the interviewer (Figure 6a). Then, on “*so it's very tight*” (Figure 6b), the hands approach each other a little with the fingertips pointing upward. Finally, on “*that means that it's a vertical gallery*” (Figure 6c), the hands form a new configuration: The gallery's height is conveyed by the distance spanned by the right hand, with its open palm turned downward at hip level (thus representing the gallery's foundation) and the left hand, also with the open palm facing down but located a little above head level (thus representing the top of the building).

- (2) *It's a twenty-five foot wide slot ... so it's very tight ... that means that it's a vertical gallery.*

Foster here adopts a frontal perspective on the gallery building, whereby his stepwise gestural construal of the gallery's physical characteristics is, compared to the factual information provided verbally, rather subjective and approximative as far as the proportions are concerned. The building's iconic representation is indexically conditioned as follows: the hands not only iconically represent the borders of the gallery but also exhibit incorporated indices pointing in three

different directions, each time creating a differently orientated bounded region of space. Furthermore, in view of the distance spanned by the hands in Figure 6b, the gesture does not exactly come across as an intuitive rendition of the attribute “*very tight*”, which could also be portrayed by bringing the hands much more closely together. This is probably because if his hands were representing the gallery’s outer walls, Foster’s body would still fit within its limits. Finally, the distance between his hands in the last gesture evoking the gallery’s height (Figure 6c) is not much greater than in the first two gestures indicating its width (Figure 6a, b). The gallery’s narrow gestalt is not really recognizable from these gestures (Figure 7 shows a photograph of the Sperone Westwater Gallery). Rather, this sequence exemplifies the viewpointed, subjective, and relative nature of gestural portrayals of this kind: they are relative with respect to the speaker’s body and the bounds of his gesture space, and, in this particular case, probably also with regard to the much wider museum buildings that Foster has previously not only experienced but also designed. All in all, his multimodal description of the gallery’s spatial dimensions reflects an external, non-immersed vantage point.



Figure 7: Sperone Westwater
Gallery, Manhattan, designed by
Norman Foster.

3.3 Immersed viewpoint construction from the inside of spatial artifacts

In the sequence captured in Figure 8 (also taken from ArchRecordTV, January 5, 2011), Foster enacts a fictive scene inside the Sperone Westwater Gallery. Immersing himself inside his mental representation of the gallery’s lobby, he depicts how one might be taken by surprise when entering the building: he mimics looking up and pointing at the underside of the massive elevator situated above the reception area. His facial expressions show that he is visibly amused (emotional viewpoint; Sweetser 2013); in (3) he simultaneously explains



Figure 8: Immersed viewpoint construction from the inside of the gallery space (“underneath the elevator”).

this special effect by employing the second-person pronoun ‘you’, thus inviting his interlocutors to virtually position themselves on this very spot inside the gallery: “*... because the last place you think you’d ever really want to be in any building is underneath the elevator*”. This instance of a *body action image icon* (Mittelberg and Waugh 2014) metonymically portrays only a few essential aspects of the full action to which it alludes; several incorporated indices further guide the viewers’ attention upward to the imagined elevator. Compared to the architecture student’s description of how she experienced a vertically extending space (shown in Figures 1–5), Foster’s portrayal exhibits both a different strategy of immersion and a higher degree of iconicity, for he simulates the given experience by involving almost his entire body: his head, eye gaze, and fingertips are all pointing upward. He thus *performs* for his audience an act of perspective-taking.

- (3) *... because the last place you think you’d ever really want to be in any building is underneath the elevator.*

What serves as a backbone for this multimodally achieved, immersive construction of imagined experiential viewpoint (Sweetser 2013) are aesthetic spatial structures with which visitors interact, instead of a larger narrative structure with characters. One possible way of experiencing this spectacular space is enacted here by the person who built the space in the first place. It is likely that Foster has observed visitors react in the surprised and amused way he imitates in his description. Or he intended this performance to make people want to visit the gallery and thus share this special experience. Compared to the constructs

character and observer viewpoint, simulated artifact immersion seems to capture the experientially and imaginatively motivated construal of these kinds of viewpoints and, thus, indexically conditioned performances more fittingly.

4 Multimodal descriptions of simulated immersion in architectural and pictorial spaces

This section explores different strategies and degrees of simulated artifact immersion observed in multimodal discourse data collected in two lab studies. The main purpose is to determine when the participants take an external perspective on the spatial artifacts they are talking about and when they resort to an immersed viewpoint to (re-)experience virtual spatial structures from within. Participants were asked to describe real-world spatial artifacts that they had previously visited (Section 4.1), their own architectural models (Section 4.2), or paintings shown to them in the lab (Section 4.3).²

In particular, we are interested in how speakers perspectivize and also move within spatial artifacts in their minds, and how this perspective-taking is enacted physically through postures and gestures while they are communicating their ideas as they dynamically evolve. Special attention will be paid to how linguistic deictic expressions, as well as indexical dimensions inscribed in the physical articulators and movements, aid in reconstructing the speakers' physical and conceptual perspectives, and how these indices anchor more general processes of immersion, imagination, and simulation in the cognitive and pragmatic contexts in question.

4.1 The body becomes part of a real-world architectural artifact

We start with a bodily portrayal exhibiting a high degree of immersion. In the following example (Figure 9a; adapted from Mittelberg et al. 2017), an architecture student evokes self-selected characteristics of the Cultural Institute of Stockholm (designed by Giò Ponti; see Figure 9b). Having previously visited and studied the building very carefully, the student decided not to use speech

² All studies were conducted in the Natural Media Lab (HumTec, RWTH Aachen University). The studies discussed in Sections 2.2, 4.1, and 4.2 were done in cooperation with the Dept. of Visual Arts (Architecture Faculty, RWTH Aachen University); for details regarding the study design and methods see Mittelberg et al. (2017).



Figure 9: (a) Simulated artifact immersion (silent) (window and light beams).
(b) Hallway, Cultural Institute of Stockholm.

in his description, which intensified the sense of immersion and inward contemplation. As shown in Figure 9a, his upward oriented left hand can be taken to represent a window in the hallway's ceiling, and his arms embody the reflections of the light beams falling into the hallway from above (see Müller 2014 on gestural modes of representation). By staying immobile like this for a short while, his portrayal effectively mediates the tranquil quality of the atmosphere reigning in the hallway. The perceptually salient, physical presence of the speaker's upward pointing arms and hands heightens the degree of iconicity that this enactment exhibits (Mittelberg 2014). His gaze is also pointing upwards in the direction of his left hand. Considering that the speaker's body is indexically anchored in real space and in the imagined space inside the building, this is an example of a real space blend (Dudis 2004; Parrill 2009; Sweetser 2012). It is also a gesturally triggered instantiation of imagined experiential viewpoint – here without any interacting linguistic space builders or supporting discourse structures.

A similar, highly immersive strategy applies when speakers do not partition their body (Dudis 2004; Sweetser 2013), that is, only use “body segments” (Calbris 1990: 44), but employ their whole body to represent an artifact, such as a spatial structure, or an object, and/or enact its behavior as fully as possible. Bouvet (1997: 17), for instance, describes a little boy pretending to be a helicopter in action by rotating his arms around the vertical axis of his body.³ These processes rely on metonymic modes in that they underpin both

³ For a continuum of different degrees of immersion in architectural design processes, see Mittelberg et al. (2017).

abstraction with respect to the gesturer and pragmatic inferencing with respect to the viewer (Mittelberg 2017).

4.2 Simulated immersion in a virtual architectural model: Social and multisensory dimensions

In the following sequences, imagined experiential viewpoint in the context of simulated artifact immersion is a constitutive factor in how the architecture student shown in Figures 10–13 conceptualizes and describes her own architectural ideas early in the design process. Although she had already visited the building site, she had not yet physically experienced the building still to be



Figure 10: (a-c) Gestures portray the speaker's imagined walking towards a building site.

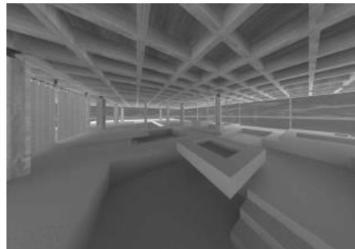


Figure 11: Preliminary architectural model of the spa.



Figure 12: Immersed viewpoint: inside the spa.



Figure 13: Simulated artifact immersion: social and multisensory dimensions.
(a) Speaker imagining seeing friends she is talking to. **(b)** Speaker pretending to look to the left and outside. **(c)** Speaker evoking the rather dim atmosphere inside the spa.

constructed there. Here, virtual models and ideas still evolving in her mind form the basis of viewpointed instances of multimodal meaning construction (for relations between gesture, imagination, and architecture, see also, e.g., LeBaron and Streeck 2000; Murphy 2005).

While the speaker is talking, in her native language German, about her design concept of a spa building and the landscape in which it is to be set, she shifts her perspective from viewing the imagined model from the outside to being situated inside it. She begins her multimodal construal with a space-building statement in (4) (“*also ich stehe vor dem Gebäude*” [so I am standing in front of the building], cf. Figure 10a) and describes how, in her imagination, she is approaching the building site (“*und gehe runter ... also ich versinke langsam*” [I am walking down ... so I am slowly sinking], cf. Figures 10b, c). Gesturally portraying how she is walking towards the spa building, she slowly moves her hands forwards while lowering them from head level to hip level. Her eye gaze is also continually directed downwards, indicating her being immersed in imagined perception and mental simulation.

- (4) *Also ich stehe vor dem Gebäude und gehe runter ... also ich versinke langsam.. und mit jeder Stufe ... verliert sich die Landschaft immer mehr über mir.. also die Weite wird immer enger.. und ich ich habe das Gefühl, dass ich dass ich in einen Ort trete ... oder dass ich jetzt irgendwie eine schützende Hülle um mich habe.*

‘So I am standing in front of the building and walking down ... so I am slowly sinking.. and with each step ... the landscape disappears more and more above me ... so the wide view gets narrower and narrower ... and I have the feeling that I am that I am entering a place ... or that I am now somehow surrounded by a protective shell.’

Next the speaker imagines herself in a specific room inside the spa building (Figure 11 shows a preliminary digital model): In (5) she says: “*ich bin quasi gerade in einem Wasserbecken*” [I am, as it were, right now in a small pool] while using her hands to indicate its approximate size and orientation (Figure 12). Note that, in reality, she remains standing in the same position during the entire recording, regardless of whether she is imagining being outside or inside the virtual spatial structures that she is describing. At every moment of interest here, she gestures with both hands in a symmetrical fashion.

- (5) *Also ich stell mir jetzt vor, dass ich (ähm) ... in diesem Badehaus bin und ich bin quasi gerade in einem Wasserbecken und ... (ähm) sitz da vielleicht mit Freunden oder mit der Familie drin.*

‘So I’m imagining now that I (uhm) ... am in this bathhouse and I am, as it were, right now in a small pool and ... (uhm) ... am sitting in there perhaps with friends or with the family.’

This sequence exhibits a multidimensional instantiation of imagined experiential viewpoint in that the speaker simulates perceiving the views, the light conditions, and the overall atmosphere from an immersed, model-internal vantage point. Combining several modalities of expression, notably, speech, gaze, and gestures, she shares her viewpointed impressions of the space’s spatial, perceptual, and social facets. In (6), she continues to interact with the virtual environment by looking ahead, to the left and right, thereby paying particular attention not only to the place’s ambience but also to possible social actors, for example, by imagining talking to friends there (Figure 13a-c).

- (6) *Ich stell mir grad vor, dass ich ... rechts seh’ ich halt irgendwie Freunde also ... mit denen ich rede, und wenn ich nach links gucke, dann hab ich (ähm) guck ich nach draußen, also es is vielleicht innen im inneren Raum auch eher so ne düstere Stimmung also eher so’n bisschen dunkler und das Licht kommt also von diesem- von dem Blick nach draußen ... (ähm) wo ich quasi ne Weite hab’ ...*
‘Right now I am imagining that I ... on the right I just somehow see friends so ... to whom I am talking, and when I look to the left, then I have (uhm) am looking outside, so there is perhaps inside inside the room a rather kind of dim atmosphere so rather kind of a bit darker and so the light is coming from this- from this view outside ... (uhm) where I have a wide view, as it were ... ’

The fact that she uses the verbs of perception *sehen* ‘see’ and *gucken* ‘look’, and frames each of the descriptions in (5) and (6) with the hint that she is ‘imaging’

the scenes while speaking, reinforces the idea of viewpointed imagined perception (Sweetser 2013). Moreover, she uses the present tense and the first-person pronoun throughout the description, which clearly enhances the sense of immersion in this particular instance of imagined experiential viewpoint construction.

An important question arises from studying these particular semiotic processes: in what ways can immersed enactments of this kind be said to be iconic? The speaker is not imitating someone else (e.g., through adopting character viewpoint), but first she is depicting herself approaching the building site via iconic traces (Figure 10) and then pretending to be sitting inside a virtual pool (Figure 12). Perhaps we can say that she is ‘being’ herself, taking on the role of an imagined experiencer guiding us, from her imagined point of view, through her virtual model and multisensorial imagery, while, intermittently, drawing our attention to her own simulated or otherwise portrayed actions within the immersed spaces; she thus becomes an icon of how she imagines herself acting in the virtual scenarios she is creating (Mittelberg 2014). While appearing to be transitioning from previously experienced outdoor places to virtual interior spaces, her body anchors this viewpointed performance in real space, in the imagined space, and in the narration space (e.g., Parrill 2012; Sweetser 2012; 2013).

Indexicality manifests itself in her postures and gestures in various ways, and these incorporated indices seem instrumental in her physical enactments of imagined experiential viewpoint. Here, the following linguistic and physical indices jointly function as grounding and experience-guiding devices: a) the verb forms in the present tense and deictic expressions, such as the first-person pronoun, spatial prepositions, and locative adverbs; b) the speaker’s eye gaze and gestures, which sometimes point in different directions but tend to be aligned in the same direction.

These observable immersive processes of simulated orientation and online imagination not only aid the speaker in advancing and communicating her ideas; the listener-observer is also invited to build up her/his own mental representation of the space, drawing together the speaker’s verbal descriptions, metonymic gestural insights, and pointers into the evolving spatial structures along with their material, social, and atmospheric qualities (e.g., Bergen 2005; Gibbs 2006).

4.3 Simulated artwork immersion: 2D paintings

Turning to the final example, we are especially interested in transmodal dimensions of simulated artifact immersion. The artifacts in question here consist of paintings, hence, they present solid two-dimensional pictorial spaces. The term *transmodal* originates in Krois’ (2011: 218) ideas on the

connection between embodiment and enactivism in the visual arts, and it particularly relates to his claim that images are not simply visual but transmodal phenomena for the beholder (see also Bredekamp 2010). This account resonates with the view on gesture taken here: gestures are not simply visual either. Gestures and postures are physical actions and viewpointed enactments that may serve not only to communicate something to somebody else so that they can follow the speaker and feel what s/he is expressing (Johnson 1987; 2007). They also serve to better understand and to re-experience one's own past physical, affective, and aesthetic experiences as well as to imagine new, evolving ideas, as we have witnessed throughout this paper (see also, e.g., Cienki and Mittelberg 2013; Gibbs 2006; McNeill 2000; Müller and Ladewig 2013).

Building on a previous study (Mittelberg 2013b), here we will consider just one of the three paintings by Paul Klee that were presented to three study participants, who were standing two steps away from a large computer screen on which each painting was displayed for one minute. Afterwards, the participants, all native speakers of American English, were asked to describe the images from memory while standing on the same spot where they had contemplated them. All three images feature a human-like figure trying to maintain a balanced posture, despite disequilibrating forces seemingly acting on their bodies.⁴ Klee's *Tightrope Walker* shows a figure that is trying to stay balanced while traversing a high wire suspended in the air (Figure 14). In their

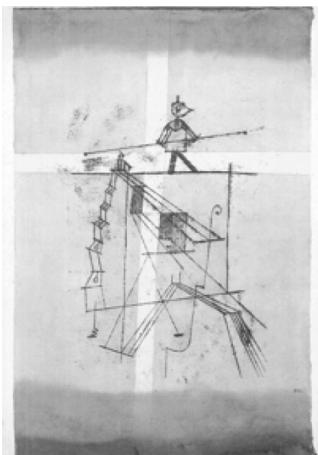


Figure 14: Paul Klee, *Tightrope Walker* (1923).

⁴ For an image schema analysis of all three paintings and the corresponding multimodal descriptions, see Mittelberg (2013b).

descriptions, all three participants systematically performed viewpoint shifts (Stec 2012) when – after detailing the pictorial composition and background of the painting from a non-immersed viewpoint – they started to enact, that is, simulate, the figure's posture and actions with their gestures, and often with their entire bodies, from an artifact-immersed viewpoint. From such image-internal vantage points, the participants also sometimes elaborated on what the figure might be thinking and sensing.

The participant shown in Figure 15 first adopts a non-immersed viewpoint when sketching the white cross and tightrope in the air with her thumb and index finger and then places an imaginary figure on top of the gesturally evoked tightrope (Figure 16 shows a motion-capture plot with the isolated traces evoking the cross and tightrope from a frontal view). Then, in mid-sentence, she shifts to an artifact-immersed viewpoint from where she simulates the figure's



Figure 15: Simulating action of a figure in a painting (immersed viewpoint).

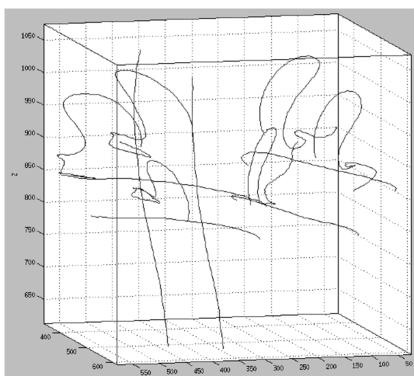


Figure 16: Background features (non-immersed viewpoint).

posture, leg position, and the action of holding a balancing pole (Figure 15). She finally portrays the figure's pointed face by seemingly extending her own nose. Interestingly, at this point, the speaker herself addresses questions of perspective, wondering whether she should turn her head to the right or to the left. She is thus taking into account the location of the interviewer, who is sitting next to the computer screen opposite her, and who she then faces in the very same way that she was facing the painting beforehand. Note that in (7) she hesitates before beginning her description with "*if you were looking at it*". When she realizes that the interviewer would have to turn around to do so, she decides to imitate the figure's posture as if she were mirroring it and says "*if I were mirroring him*". Except for this instance, she uses the third-person pronoun throughout her verbal description.

- (7) *Uhm, in the ... background of the picture there is like a white cross ... uhm, which is very different from the rest of the picture which is also ... uhm, like tan except for the bottom where it could (...) it almost looks like it could be green ... uhm, and then, on top of that, it looks like ... uhm ... like a man or some sort of figurine is standing on it ... (VIEWPOINT SHIFT), and he's holding something and he ... is looking to the side, so it almost looks like ... he's going like this, 'cos one leg looks, uhm ... more stretched out than the other ... and ... his uhm, head is drawn and the face are like a teardrop, so ... it's very ... pointed.. and ... uhm he's like this ... if you're looking at ... if I were mirroring him, umh, and so from this side ...*

Mentally picturing and describing a static painted scene, which features a figure traversing a path, tends to involve fictive motion (Matlock 2004; 2006; Talmy 2000), especially when speakers are not only mimicking a figure's static pose but also incorporating dynamic aspects into the scenario by moving their arms, heads, etc. (Mittelberg 2013b). Here we need to distinguish between different kinds of gestural movements, namely, between those that produce iconic sketches of the compositional structure of an image (Figure 16), those that trace the fictive motion of a figure or an entity within the image, and those that actually simulate an action or abstract process (Arnheim 1969; Mittelberg and Waugh 2014; Müller 2014; Streeck 2009).

Observant listeners, for their part, build up their mental representations based on the viewpointed, multimodal portrayals they are attending to, thus trying to imagine and follow what is being communicated to them by drawing on their own experiential knowledge, e.g., mental frames (Fillmore 1982; Mittelberg 2017), and exercising their embodied capacity for simulation and imagination (Bergen 2005; Gibbs and Matlock 2008). Gestures and whole-body

enactments like the ones discussed here thus physically transpose what Johnson (2005: 31) calls “felt qualities of meaning, experience and understanding” (see also Johnson 2007; Turner 2006).

5 Discussion: Immersion, indexicality, and metonymy

Immersive, indexical, and metonymic modes have been touched upon throughout the present paper; they will now serve as anchor points for the ensuing discussion. The remarks offered in the following sections can only be preliminary at this point, as these complex processes of multimodal viewpoint and meaning construction still need to be worked out in more detail and examined on the basis of more data.

5.1 Immersion: On the inside of spaces, events and imaginative processes

As spatial artifacts constitute the core interest in this paper, one main goal was to find a way of accounting for the immersive, imaginative, and multimodal dimensions of how people perceive and describe them. The varying degrees and strategies of simulated artifact immersion that manifested themselves in the speakers’ multimodal discourses seem to be motivated by the fact that there were no preexisting discourse structures, with a story line and a set of characters, for the speakers to fall back on. In the case of the Klee painting (Figure 14), the concept of character viewpoint could, in principle, fit, given that there is a human-like figure that takes center stage in the image space. Yet simulated artifact immersion reinforces the transmodal nature of the experience and appreciation of such artworks (Johnson 2007; Krois 2011; Mittelberg 2013b).

The idea of immersion, in conjunction with the concept of imagined experiential viewpoint (Sweetser 2013), has the further advantage of highlighting not only spatial but also temporal dimensions of the embodied cognitive processes involved both with respect to the speaker-gesturer and to the listener-observer. As for the narrations examined, the architecture student who spoke about a childhood memory used only the past tense in all of her utterances (Section 2.2). By contrast, the other architecture student consistently used the present tense to talk about her spa design (Section 4.2), which highlights the fact that the design process was still in progress. In the English description of the Klee painting, the present tense was used to detail the background features, whereas the figure’s

actions were described in the present progressive (e.g., *is standing, holding, looking, going like ...*) (Section 4.3). When witnessing the ongoing processes of imagining, sensing, moving, gesturing, and incremental utterance formation in the video data, one indeed feels drawn into the scenarios described. This is reminiscent of similar construal effects prompted by the use of the progressive in English written texts and the use of imperfective verb forms in general: They make us experience ongoing situations from the inside (e.g., Dancygier 2012; Sweetser 2013; Vandelanotte 2009).

5.2 Indexicality: Experiential, spatial, and cross-modal anchorage

Indexicality is inherent in the communicating body and it is generally one of the strong driving forces motivating multimodal pragmatics (e.g., Mittelberg 2017; Sweetser 2012). Without incorporating indexical elements in their speech, gaze, and bodily communicative action, interlocutors cannot perspectivize their (iconic) portrayals by setting up physical and interactive conversational contexts (Jakobson 1971). Nor could the participants of the studies reported on here have simultaneously positioned themselves and their descriptions in the face-to-face exchanges in the lab space *and* in their imaginary worlds, that is, in blended mental spaces (e.g., Dudis 2004; Parrill 2012; Sweetser and Stec 2016).

A related concept of relevance is that of *material anchors* (Hutchins 2005). Material structures and artifacts have been shown to serve as physically present or imaginary material anchors for cognitive operations and discursive, interactive processes (Coulson 2001; Dancygier 2012; Fauconnier and Turner 2002). Sweetser (2012) sees the human body itself as a powerfully functioning material anchor that naturally has a viewpoint:

Gesture, like language, shows deictic centers, displacement phenomena and blended deictic structure. Our bodies are the most flexible and powerful material anchors [...] for representing and expressing viewpoint. [...] there is no more powerful icon for a bodily viewpoint than an actual body with an actual inherent viewpoint. Gestures therefore express spatial indexicality via embodied spatial indexicality – for humans, a pointing gesture directs joint attention to a particular actual location relative to the actual pointing body. This in itself is a cognitive achievement. (Sweetser 2012: 13)

This citation captures some of the main insights that we can draw from the theoretical issues and the multimodal discourse segments discussed in this paper. In the analyzed segments, the speakers' bodies all serve as permanent, visible anchors throughout their explanations, thus flexibly integrating real space and the imagined spaces and spatial artifacts in which they immerse themselves in

their imagination. They stay with their indexically anchored bodies *in situ*, while their heads, arms, and hands point to imagined objects and locations. If we compare, for instance, how the first speaker portrays her way to kindergarten (in Section 2.2) and the fourth speaker leads us to the spa building site (in Section 4.2), they both adopt a similar semiotic practice: their bodies remain in the same place while they are moving their hands away from their torsos to trace paths of their own movement in the imagined scenarios.

We also saw how, together with gestural viewpoint markers, linguistic indexical elements – such as verb tenses, pronouns, as well as locative and temporal prepositions and adverbs – constitute a dynamic multimodal fabric that allows the analyst to reconstruct the vantage points from where an imagined scenario becomes incrementally set up and perspectivized.

5.3 Metonymy: Partial perception, partial bodies, and partial portrayal

Human perception is not only inherently viewpointed (Sweetser 2013), it is also inherently partial, that is, metonymic. Gestures, too, are by their nature metonymic in one way or another (e.g., Mittelberg 2006; Mittelberg and Waugh 2014; Müller 1998). Body partitioning in gesture and sign language, as introduced by Dudis (2004), also seems to be rooted in metonymic principles and relations. In the present paper, the gestural figurations and indications of what the speakers were remembering or newly conceiving exhibited a high degree of partiality and schematicity with regard to which parts of the respective spatial structures or bodily actions were being profiled. In the moment of gestural expression, many aspects are necessarily abstracted or omitted, whether due to partial perception, partial memory, or partial motivating conceptual structures (Mittelberg 2017). With their gestures and postures, speakers may convey fragmentary and schematic, yet essential qualities of their experiences and mental representations, as the architecture student (in Section 2.2) did when she was drawing with her index finger, minimally but effectively, representing herself as an image-schematic spiral gestalt on her way to kindergarten in the air (Figures 2 and 3), or when Foster was simulating a surprised reaction inside the gallery's lobby in a metonymically reduced fashion (Figure 8).

The principles behind the tight interaction of viewpoint and metonymy, as a function of the artifacts and specific discourse contexts involved, still need to be worked out in more detail. One way of going about this could be to see how speakers draw on different kinds of contiguity relations available in their physical and social environment as well as in the imagined scenarios they are talking about. As shown by previous research across semiotic systems, different

kinds of contiguity may be operationalized in meaning-making processes (Jakobson and Pomorska 1983). One may draw on inner contiguity, that is, part-whole relations internal to an entity, structure, or action (giving rise to internal metonymy), or on outer contiguity relations, such as contact or adjacency, between a hand and an imaginary object, surface, or instrument that the gesturer pretends to be manipulating (giving rise to external metonymy; for more details see Mittelberg and Waugh 2009, 2014). In this account, ‘internal’ means internal to a structure, an object, or an action routine to which a gesture alludes by singling out relevant features (Arnheim 1969). This entity or action does not necessarily have to be construed from a scene-internal viewpoint, as we have seen with the example of the cross-like background structure isolated from the Klee painting *Tightrope Walker* (Figures 14 and 16) and Foster’s portrayals of the gallery’s vertical gestalt (Figure 6). Likewise, an outer contiguity relation may become meaningful in gestures reflecting a construal from an artifact-immersed viewpoint. For instance, the speaker imitating the figure in the same painting seems to be holding the balancing pole alluded to in her verbal utterance (Figure 15). Drawing on cross-modal mechanisms of pragmatic inferencing (Mittelberg 2017), the pole can be metonymically inferred from the simulated action of holding it; this is thus an example of external metonymy – or frame metonymy (Dancygier and Sweetser 2014; Taub 2001) – performed from an immersed, image-internal vantage point.

Given these different, experientially motivated principles that feed into gestural sign constitution, it seems particularly promising to further investigate how the immersed viewpoint strategies discussed in this paper pattern with distinct metonymic principles and modes of representation (Müller 1998; 2014) across various artifacts, languages, and genres (Deliema and Sweetser 2016; Mittelberg and Waugh 2014: 1762).

6 Concluding remarks

The chief goal of this paper was to introduce simulated artifact immersion as a viewpoint strategy that is typically instantiated multimodally when speakers are talking about their experiences with spatial artifacts. Building on the notion of imagined experiential viewpoint, it was argued that simulated artifact immersion can be seen as an alternative concept to (combinations of) character and observer viewpoint in semiotic contexts where this distinction does not seem to hold or do justice to the cognitive-communicative task, artifact, and discourse at hand. Furthermore, it was shown that the viewpointed nature of multimodal descriptions owes a great deal to their inherent indexicality. The fact that

imagining spatial scenarios and simulating actions takes time was also emphasized. How temporal dimensions of spatial artifact immersion factor into aspectual framing and the use of verb tenses still needs to be more systematically accounted for in future gesture research.

A century ago, Saussure (1986: 23) made the observation that it is not the object which precedes the point of view but the point of view which creates the object.⁵ In this paper, we saw how spatial artifacts are (re-)created from different viewpoints through linguistic and gestural means. In the process, perception and semiotic representation both result, in their own specific ways, in viewpointed, partial (metonymic) insights into locally, or prototypically, salient aspects of an artifact, an ongoing discourse, and otherwise relevant elements in the cognitive, physical, and socio-cultural context. Since the viewpoint(s) adopted seem to have an influence on what kind of metonymic modes may be operationalized and which gestural practices are chosen, future research is needed to illuminate how exactly this influence plays out in spoken and signed multimodal usage events. In this connection, the interaction of viewpoint, metaphor, and gesture also seems to be worth investigating with regard to abstract artifacts, knowledge domains, and conceptual structures.

All in all, it seems indeed worthwhile to keep working towards a more complete picture of how the viewpoint strategies and construal operations addressed above play into both the production and interpretation of multimodal discourse. While the multidisciplinary research discussed here shows considerable progress, more empirical studies on how viewpoint phenomena function across languages, modalities, discourse genres, and contexts of use are required to further narrow in on the general embodied principles that underpin multimodal meaning construction.

Acknowledgements: I would like to thank Barbara Dancygier and Lieven Vandelanotte for their valuable input and for including my work in this special issue on viewpoint phenomena in multimodal communication. I also thank John Newman and two anonymous reviewers for their insightful comments and suggestions. Special thanks also go to Hannah Groninger and Thomas Schmitz for cooperating on the architectural design studies, to Bela Brenger and Matthias Priesters for creating the motion-capture plots, and to Mary Copple for helpful feedback on the manuscript.

Funding: Excellence Initiative of the German Federal and State Governments.

⁵ Saussure (1986: 23): “Bien loin que l’objet précède le point de vue, on dirait que c’est le point de vue qui crée l’objet.”

References

- Arnheim, R. 1969. *Visual thinking*. Berkeley, CA: University of California Press.
- Bergen, Benjamin K. 2005. Mental simulation in literal and figurative language understanding. In Seana Coulson & Barbara Lewandowska-Tomaszczyk (eds.), *The literal and the non-literal in language and thought*, 255–278. Frankfurt a. M: Peter Lang.
- Bergen, Benjamin K. 2012. *Louder than words: The new science of how the mind makes meaning*. New York: Basic Books.
- Bouvet, Danielle. 1997. *Le corps et la métaphore dans les langues gestuelles: A la recherche des modes de production des signes*. Paris: L'Harmattan.
- Bredenkamp, Horst. 2010. *Theorie des Bildakts*. Frankfurt a.m.: Suhrkamp.
- Brône, Geert, Bert Oben, Annelies Jehoul, Jelena Vranjes & Kurt Feyaerts. 2017. Eye gaze and viewpoint in multimodal interaction management. DOI:10.1515/COGL-2016-0119.
- Calbris, Geneviève. 1990. *The semiotics of French gestures*. Cambridge: Cambridge University Press.
- Cienki, Alan. 2013. Cognitive linguistics: Spoken language and gesture as expressions of conceptualization. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Teßendorf (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.1), 182–201. Berlin & Boston: De Gruyter Mouton.
- Cienki, Alan. 2015. Spoken language usage events. *Language and Cognition* 7(4). 499–514.
- Cienki, Alan & Irene Mittelberg. 2013. Creativity in the forms and functions of gestures with speech. In Tony Veale, Kurt Feyaerts & Charles Forceville (eds.), *Creativity and the agile mind: A multi-disciplinary study of a multi-faceted phenomenon*, 231–252. Berlin & New York: Mouton de Gruyter.
- Clark, Herbert H. 2016. Depicting as a method of communication. *Psychological Review* 123(3). 324–347.
- Coulson, Seana. 2001. *Semantic leaps*. Cambridge: Cambridge University Press.
- Croft, William & Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Dancygier, Barbara. 2012. *The language of stories: A cognitive approach*. Cambridge: Cambridge University Press.
- Dancygier, Barbara & Eve Sweetser (eds.) 2012. *Viewpoint in language: A multimodal perspective*. Cambridge: Cambridge University Press.
- Dancygier, Barbara & Eve Sweetser (eds.) 2014. *Figurative language*. Cambridge: Cambridge University Press.
- Dancygier, Barbara, Lu Wei-Lun & Arie Verhagen (eds.) 2016. *Viewpoint and the fabric of meaning: Form and use of viewpoint tools across languages and modalities*. Berlin: Mouton de Gruyter.
- Debras, Camille. 2013. *Multimodal stancetaking in a videotaped corpus of discussions about the environment in British English*. Paris: Sorbonne Nouvelle – Paris 3 University dissertation.
- Deliema, David & Eve Sweetser. 2016. Rethinking gestural viewpoint as multidimensional rather than a dichotomy. Conference talk at the International Society for Gesture Studies Conference (ISGS7), Paris, France, 18–22 July.
- Dudis, Paul. 2004. Body partitioning and real-space blends. *Cognitive Linguistics* 15(2). 223–238.

- Duncan, Susan D., Justine Cassell & Elena T. Levy (eds.) 2007. *Gesture and the dynamic dimension of language*. Amsterdam & Philadelphia: John Benjamins.
- Ehmer, Oliver. 2011. *Imagination und Animation. Die Herstellung mentaler Räume durch animierte Rede*. Berlin & New York: Mouton de Gruyter.
- Emmorey, Karen, Barbara Tversky & Holly A. Taylor. 2001. Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation* 2. 157–180.
- Fauconnier, Gilles & Mark Turner. 2002. *The way we think*. New York: Basic Books.
- Fillmore, Charles J. 1982. Frame semantics. In The Linguistic Society of Korea (eds.), *Linguistics in the morning calm*, 111–137. Seoul: Hanshin.
- Gibbs, Raymond W., Jr. 2006. *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
- Gibbs, Raymond W., Jr. & Teenie Matlock. 2008. Metaphor, imagination, and simulation: Psycholinguistic evidence. In Raymond W. Gibbs Jr. (ed.), *The Cambridge handbook of metaphor and thought*, 161–176. Cambridge: Cambridge University Press.
- Hostetter, Autumn B. & Martha Alibali. 2008. Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review* 15(3). 495–514.
- Hutchins, Edwin. 2005. Material anchors for conceptual blends. *Journal of Pragmatics* 37. 1555–1577.
- Jakobson, Roman. 1971. Shifters, verbal categories, and the Russian verb. In Roman Jakobson (ed.), *Selected writings, Vol. II: Words and language*, 130–147. The Hague & Paris: Mouton.
- Jakobson, Roman & Krystyna Pomorska. 1983. *Dialogues*. Cambridge, MA: MIT Press.
- Janzen, Terry. 2012. Two ways of conceptualizing space: Motivating the use of static and rotated vantage point space in ASL. In Barbara Dancygier & Eve E. Sweetser (eds.), *Viewpoint in language: A multimodal perspective*, 156–176. Cambridge: Cambridge University Press.
- Johnson, Mark. 1987. *The body in the mind: The bodily basis of meaning, imagination and reasoning*. Chicago: Chicago University Press.
- Johnson, Mark. 2005. The philosophical significance of image schemas. In Beate Hampe (ed.), *From perception to meaning: Image schemas in cognitive linguistics*, 15–33. Berlin: Mouton de Gruyter.
- Johnson, Mark. 2007. *The meaning of the body: Aesthetics of human understanding*. Chicago: University of Chicago Press.
- Krois, John M. 2011. *Körperbilder und Bildschemata. Aufsätze zur Verkörperungstheorie ikonischer Formen* (Actus et imago 2), hrsg. v. Horst Bredekamp & Marion Lauschke. Berlin: Akademie Verlag.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar, Vol.1, Theoretical perspectives*. Stanford: Stanford University Press.
- LeBaron, Curtis & Jürgen Streeck. 2000. Gestures, knowledge and the world. In David McNeill (ed.), *Language and gesture*, 118–138. Cambridge: Cambridge University Press.
- Liddell, Scott. 2003. *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.
- MacWhinney, Brian. 2005. The emergence of grammar from perspective. In Diane Pecher & Rolf Zwaan (eds.), *Grounding cognition*, 198–223. Cambridge: Cambridge University Press.
- Marghetis, Tyler & Benjamin K. Bergen. 2014. Embodied meaning inside and out: The coupling of gestures and mental simulation. In Cornelia Müller, Alan Cienki, Ellen Fricke,

- Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.2), 2000–2007. Berlin & Boston: De Gruyter Mouton.
- Matlock, Tenie. 2004. Fictive motion as cognitive simulation. *Memory & Cognition* 32. 1389–1400.
- Matlock, Tenie. 2006. Depicting fictive motion in drawings. In June Luchjenbroers (ed.), *Cognitive linguistic investigations: Across languages, fields, and philosophical boundaries*, 67–85. Amsterdam & Philadelphia: John Benjamins.
- Matlock, Tenie, David Sparks, Justin L. Matthews, Jeremy Hunter & Stephanie Huette. 2012. Smashing new results on aspectual framing: How people talk about car accidents. *Studies in Language* 36(3). 699–720.
- McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, David (ed.) 2000. *Language and gesture*. Cambridge: Cambridge University Press.
- Mittelberg, Irene. 2006. *Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar*. Ann Arbor, MI: UMI Cornell University dissertation.
- Mittelberg, Irene. 2013a. The exbodyed mind: Cognitive-semiotic principles as motivating forces in gesture. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Teßendorf (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.1), 750–779. Berlin & Boston: de Gruyter Mouton.
- Mittelberg, Irene. 2013b. Balancing acts: Image schemas and force dynamics as experiential essence in pictures by Paul Klee and their gestural enactments. In Mike Borkent, Barbara Dancygier & Jennifer Hinnell (eds.), *Language and the creative mind*, 325–346. Stanford: CSLI Publications.
- Mittelberg, Irene. 2014. Gestures and iconicity. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.2), 1712–1732. Berlin & Boston: De Gruyter Mouton.
- Mittelberg, Irene. 2017. Embodied frames, metonymy and pragmatic inferencing in gesture. *Gesture* 16(2).
- Mittelberg, Irene, Thomas H. Schmitz & Hannah Groninger. 2017. Operative manufacts: Gestures as embodied sketches in early stages of the design process. In Sabine Ammon & Remei Capdevila Werning (eds.), *The active image: Architecture and engineering in the age of modeling*, 99–131. Berlin: Springer.
- Mittelberg, Irene & Linda R. Waugh. 2009. Metonymy first, metaphor second: A cognitive-semiotic approach to multimodal figures of thought in co-speech gesture. In Charles Forceville and Eduardo Urios-Aparisi (eds.), *Multimodal metaphor*, 329–356. Berlin & New York: Mouton de Gruyter.
- Mittelberg, Irene & Linda R. Waugh. 2014. Gestures and metonymy. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.2), 1747–1766. Berlin & Boston: De Gruyter Mouton.

- Müller, Cornelia. 1998. *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berlin Verlag Spitz.
- Müller, Cornelia. 2014. Gestural modes of representation as techniques of depiction. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Jana Bressem (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.2), 1687–1702. Berlin & Boston: De Gruyter Mouton.
- Müller, Cornelia & Silva H. Ladewig. 2013. Metaphors for sensorimotor experiences: Gestures as embodied and dynamic conceptualizations of balance in dance lessons. In Mike Borkent, Barbara Dancygier & Jennifer Hinnell (eds.), *Language and the creative mind*, 295–324. Stanford: CSLI Publications.
- Murphy, Keith M. 2005. Collaborative imagining: The interactive use of gestures, talk, and graphic representation in architectural practice. *Semiotica* 156. 113–145.
- Parrill, Fey. 2009. Dual viewpoint gestures. *Gesture* 9(3). 271–289.
- Parrill, Fey. 2010. Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes* 25(5). 650–668.
- Parrill, Fey. 2012. Interactions between discourse status and viewpoint in co-speech gesture. In Barbara Dancygier & Eve E. Sweetser (eds.), *Viewpoint in language: A multimodal perspective*, 97–112. Cambridge: Cambridge University Press.
- Peirce, Charles S. 1960. In Charles Hartshorne & Weiss. Paul (eds.), *Collected papers of Charles Sanders Peirce (1931–1958)*. Vol. I.: *Principles of philosophy*, Vol. II: *Elements of logic*. Cambridge: The Belknap Press of Harvard University Press.
- Perlman, Markus & Raymond W. Gibbs Jr. 2013. Sensorimotor simulation in speaking, gesturing, and understanding. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill & Sedinha Teßendorf (eds.), *Body – Language – Communication. An international handbook on multimodality in human interaction* (Handbooks of Linguistics and Communication Science 38.1), 512–533. Berlin & Boston: De Gruyter Mouton.
- Saussure, Ferdinand de. 1986. In par Charles Bally, Albert Sechehaye & T. De Mauro. (eds.), *Cours de linguistique générale de Ferdinand de Saussure*. Paris: Payot.
- Schoonjans, Steven. 2014. *Modalpartikel als multimodale Konstruktionen. Eine korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen*. Leuven: KU Leuven unpublished dissertation.
- Stec, Kashmira. 2012. Meaningful shifts: A review of viewpoint markers in co-speech gesture and sign language. *Gesture* 12(3). 327–360.
- Streeck, Jürgen. 2009. *Gesturecraft: The manu-facture of meaning*. Amsterdam & Philadelphia: John Benjamins.
- Sweetser, Eve. 2007. Looking at space to study mental spaces: Co-speech gesture as a crucial data source in cognitive linguistics. In Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson & Michael Spivey (eds.), *Methods in cognitive linguistics*, 201–224. Amsterdam & Philadelphia: John Benjamins.
- Sweetser, Eve. 2012. Introduction: Viewpoint and perspective in language and gesture, from the ground down. In Barbara Dancygier & Eve E. Sweetser (eds.), *Viewpoint in language: A multimodal perspective*, 1–22. Cambridge: Cambridge University Press.
- Sweetser, Eve. 2013. Creativity across modalities in viewpoint construction. In Mike Borkent, Barbara Dancygier & Jennifer Hinnell (eds.), *Language and the creative mind*, 239–254. Stanford, CA: CSLI Publications.

- Sweetser, Eve & Kashmiri Stec. 2016. Maintaining multiple viewpoints with gaze. In Barbara Dancygier, Lu Wei-Lun & Arie Verhagen (eds.), *Viewpoint and the fabric of meaning: Form and use of viewpoint tools across languages and modalities*, 237–257. Berlin: Mouton de Gruyter.
- Talmy, Leonard. 2000. *Towards a cognitive semantics*. Cambridge, MA: MIT Press.
- Taub, Sarah. 2001. *Language from the body. Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.
- Turner, Mark (ed.) 2006. *The artful mind*. Oxford: Oxford University Press.
- Vandelanotte, Lieven. 2009. *Speech and thought representation in English: A cognitive-functional approach*. Berlin: Mouton de Gruyter.
- Verhagen, Arie. 2007. Construal and perspectivization. In Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*, 48–81. Oxford: Oxford University Press.