Viewpoint matters: Prototypical vs. non-prototypical co-speech gestures in the VP domain[1]

Sebastian WALTER — *Goethe University Frankfurt*
Cornelia EBERT — *Goethe University Frankfurt*
Stefan HINTERWIMMER — *University of Hamburg*

Abstract. The present paper extends Ebert et al.'s (2020) theory of the semantic contribution of co-speech gestures to the VP domain by investigating differences between *prototypical* and *non-prototypical* iconic co-speech gestures. Drawing from Ebert et al. (2022) and Ebert and Hinterwimmer (2022), we propose a formalism where a prototypical gesture denotes an event type which is not interpreted iconically, whereas non-prototypical gestures depend on a viewpoint variable and are interpreted iconically. Moreover, Ebert et al.'s (2020) similarity predicate SIM is spelled out for the verbal domain. We argue that viewpoint is, in the verbal domain, one factor the similarity predicate can depend on.

Keywords: gesture semantics, iconicity, perspective dependence, viewpoint.

1. Introduction

The study of co-speech gestures has gained considerable attention in formal semantics, particularly in the past decade, as researchers have developed dedicated theoretical frameworks to account for the meaning contribution of gestures within linguistic theory (e.g., Ebert and Ebert, 2014; Schlenker, 2018; Ebert et al., 2020). Co-speech gestures, which naturally accompany spoken language, provide additional information that interacts with the verbal utterance in systematic ways. Most formal semantic theories agree that the content of co-speech gestures is projective and hence not-at-issue (e.g., Ebert and Ebert, 2014; Schlenker, 2018; Ebert et al., 2020, but see Esipova, 2019 for an alternative account).

Ebert et al. (2020) argue that co-speech gestures pattern in their semantic and discourse behavior with supplements and hence analyze them as conventional implicatures in the sense of Potts (2005). Moreover, they argue that iconic gestures can be interpreted either as object-related, directly referring to an individual, or as concept-related, providing an abstract representation of a prototypical concept (Fricke, 2012). This distinction is closely tied to the temporal alignment of gesture and speech, which has been shown to systematically influence interpretation (Ebert et al., 2022). However, a shortcoming of the account by Ebert et al. (2020) is that it strongly focuses on the nominal domain. Since verbs describe events rather than objects, extending their approach to the verbal domain requires additional theoretical considerations.

In this paper, we extend the supplemental account proposed by Ebert et al. (2020) to the VP domain, thereby bridging this theoretical gap. Specifically, we introduce a distinction between *prototypical* and *non-prototypical* iconic gestures in the verbal domain. Our key claim is that prototypical gestures denote event types, abstracting away from specific real-world instances, whereas non-prototypical gestures denote event tokens and are interpreted iconically, requiring a direct resemblance to a particular event. This distinction closely parallels the object-related

[1]We would like to thank the audience of *Sinn und Bedeutung 29* for their feedback on the work presented in this paper. This research was conducted as part of the priority program 2392 *ViCom* funded by the German Research Foundation (DFG). We thankfully acknowledge their financial support.

vs. concept-related interpretations of iconic gestures observed in the nominal domain Fricke (2012), but we argue that in the VP domain, it is mediated by viewpoint dependence. That is, non-prototypical gestures depend on a viewpoint variable that determines their interpretation in relation to the event described by the accompanying speech.

To formalize this distinction, we build on the similarity predicate SIM proposed by Ebert et al. (2020), adapting it to event semantics. Our analysis predicts that viewpoint plays a crucial role in determining whether a gesture functions as a prototypical or non-prototypical representation of an event. This perspective-dependent interpretation aligns with findings from sign language research, where the semantics of classifier constructions has been argued to involve viewpoint-based iconicity (Schlenker and Lamberton, 2024).

The structure of this paper is as follows: In Section 2, we provide an overview of previous research on gesture semantics, thereby briefly sketching the supplemental account (Ebert and Ebert, 2014; Ebert et al., 2020), focusing on how co-speech gestures interact with linguistic meaning. Section 3 presents novel empirical observations on the interpretation of prototypical and non-prototypical gestures in the VP domain. In Section 4, we develop a formal analysis that incorporates event semantics into the supplemental approach, capturing the role of viewpoint in gesture interpretation. Finally, Section 5 discusses the broader implications of our findings, outlining directions for future research.

2. Gesture semantics

Gestures are a fundamental part of human communication, interacting with speech in systematic ways to convey meaning. While gestures have traditionally been studied in psychology, anthropology, and gesture studies, their integration into formal semantics is a relatively recent development. Gestures are not merely redundant or illustrative but contribute structured, interpretable meaning that interacts with spoken language at multiple levels (McNeill, 1992; Kendon, 2004; Ebert, 2024). In this paper, we are concerned with a particular type of gestures, namely iconic gestures. This type of gesture depicts certain contextually relevant aspects of its referent.

Gesture and speech are temporally aligned (McNeill, 1992; de Ruiter, 1998; Kendon, 2004; Loehr, 2004; Ebert et al., 2011). Crucially, as we will elaborate in more detail below, the temporal alignment has been shown to affect the interpretation of a speech-accompanying gesture (Ebert et al., 2022). Speech-accompanying gestures can occur pre- or post-speech, that is, either before or after the speech signal they are semantically associated with. They can also occur co-speech, meaning they co-occur with the speech unit they are associated with. This is the case we will be engaged with throughout the present paper. Finally, gestures can occur pro-speech. In this case, they replace an entire part of speech.

Gesture and speech jointly contribute meaning to an utterance (e.g., Kendon, 1980; McNeill, 1992; de Ruiter, 1998; Ebert, 2024). Although this observation has been established in fields neighboring formal linguistics, formal semanticists have only very recently discovered their interest in the type of meaning contribution gestures add to an utterance. Consider the example in (1). Small caps are used to represent gestures throughout the entire paper, square brackets and indentation are used to indicate gesture-speech alignment.

(1) I brought [a bottle of water] to the talk.
BIG

(Ebert and Ebert, 2014)

The iconic co-speech gesture that is aligned with the DP *a bottle of water* in (1) provides additional information about the bottle, i.e., it adds the information that the bottle was big.[2] As argued by Ebert et al. (2020), the meaning contribution of co-speech gestures has a special semantic status, contributing an entire proposition. This proposition, however, is backgrounded and secondary in contrast to the proposition expressed by the speech signal it co-occurs with. Ebert et al. (2020) observe that co-speech gestures share this backgroundedness with supplements, such as appositives. Therefore, they analyze them as conventional implicatures in the sense of Potts (2005). This leads them to hypothesize that they contribute so-called not-at-issue meaning by default (Potts, 2005), like appositives. Not-at-issue content is secondary content the speaker does not wish to steer the conversation toward. It is, put metaphorically, not 'on the table' for discussion (Farkas and Bruce, 2010). Not-at-issue content contrasts with so-called at-issue content, the asserted content, which is the main claim of an utterance. Not-at-issue content has been argued to project through negation and other entailment-cancelling operators.

(2) I did not bring [a bottle of water] to the talk. #A small one is enough for me.
BIG

(Ebert and Ebert, 2014)

The infelicity of the continuation in the example above illustrates what is meant by projection (Langendoen and Savin, 1971): The information contributed by the gesture in the first sentence seems to escape the scope of the negation. The size information given in the continuation contradicts the size information provided by the gesture which projected out of the negation, resulting in infelicity. In other words, projective content can be understood as content the speaker is still committed to after embedding it under an entailment-cancelling operator (Chierchia and McConnell-Ginet, 1990).

Not-at-issue content differs not only in its semantic behavior from at-issue content, but also in its discourse behavior. For example, it cannot be directly negated in discourse, as evidenced by the short dialog in (3), where B_1's attempt to deny the meaning contributed by the gesture results in infelicity:

(3) A: I brought [a bottle of water] to the talk.
BIG
B_1: #No, you only brought a small one.

What has to be done instead to target the content of the gesture, or, more generally, of not-at-issue content, is to use a discourse-interrupting element, such as *Hey, wait a minute!* (Shanon, 1976; von Fintel, 2004):

(4) B_2: Hey, wait a minute! The bottle you brought isn't that big.

Alternatively, as noted by Tonhauser (2012), one can assent with the at-issue proposition expressed by the verbal at-issue part of A's utterance in (3) followed by an adversative continuation targeting the content contributed by the gesture, as B_3 does in (5).

[2]When discussing data, we will henceforth often abbreviate *iconic co-speech gesture* to *gesture*.

(5) B_3: Yes, but the bottle actually wasn't that big.
B_4: #Yes, but you didn't bring a bottle to the talk.

Crucially, this only works for not-at-issue content. If the adversative continuation targets the at-issue content of A's utterance, this results in infelicity, which is shown by the utterance of B_4 in the example above.

Before turning to the formal proposal of Ebert et al. (2020), we would like to discuss an interesting observation by Fricke (2012) that there seem to be different interpretations of co-speech gestures. She reports the findings from a study where participants had to describe a route they had previously walked through Berlin to someone else. On this route, they passed by a gate that was of rectangular shape. Interestingly, when their description of this gate included a gesture, participants did not always produce a rectangular gesture, as one would expect. Instead, some of them produced a bow-formed gesture. This raises the question of whether the latter group produced a mismatching gesture, which is answered in the negative by Fricke (2012), since this pattern occurred too frequently and systematically to be labeled a proper mismatch. Instead, she claims that the group of participants that produced the alleged mismatch produced another type of gesture compared to the group that produced a rectangular gesture. Fricke (2012) argues that a gesture does not always refer to a memory of a certain object (i.e., to a token of a type of object). Alternatively, it can refer to an abstract mental concept, that is, a prototype or a type of an object. Gestures referring to a memorized real-world object receive a so-called *object-related interpretation* ('objektbezogene Lesart'), whereas gestures referring to an abstract mental concept receive a *concept-related interpretation* ('interpretantenbezogene Lesart'), two terms which we will adopt for gesture interpretations in the nominal domain throughout this paper.

To validate her claim that gestures can receive different interpretations, a further study was conducted by Fricke (2012). This time, participants, who were all native speakers of German, had a fairly simple task: they were instructed to draw a picture of a prototypical gate. This was done to determine what shape gates normally have for German native speakers. Out of 22 participants, 16 drew a gate which had a bow-formed shape. The remaining six did not draw a gate at all. Instead, they drew a picture of a soccer goal, which is due to the German word for *gate* ('Tor') being ambiguous between the two concepts. Crucially, what this study shows is that gates are considered to be prototypically bow-formed, at least for the German-speaking community (in Germany), thereby confirming Fricke's (2012) hypothesis that those participants who produced a bow-formed gesture really produced a concept-related gesture and not a mismatching one.

Fricke (2012) investigates the syntactic differences between the two gesture types in a next step. She argues that a concept-related gesture is strictly speaking not attributive. Hence, it is unable to semantically modify the extension of the noun it co-occurs with. Rather, this type of gesture provides intensional information about the concept they refer to. By contrast, since object-related gestures refer to memorized objects, they can function attributively. In other words, they are able to modify the extension of the core noun of an NP.

Ebert et al. (2020) incorporate the idea to distinguish between object- and concept-related interpretations of gestures into their formal system treating gestures as supplements. They depart from Fricke (2012), however, by positing that the distinction between the two interpretations

comes about by means of different alignment patterns of gesture and speech in the nominal domain. We will elaborate on this below after having sketched the core assumptions of Ebert et al.'s (2020) formal model.

Ebert et al. (2020) claim that iconic and pointing gestures essentially function very similarly: both refer to an individual, the gesture referent g. The crucial difference between the two types of gestures is that while pointing gestures pick out their referent directly, with deferred reference being possible (Nunberg, 1993), the gesture referent introduced by performing an iconic gesture is abstract. Crucially, this abstract individual carries at least the features which are crucial for comparison (Umbach and Gust, 2014). These features are determined by the context.

As mentioned above already, Ebert et al. (2020) take different alignment patterns of gesture and speech to be the decisive factor distinguishing between object- and concept-related gestures. Specifically, they identify three different alignment patterns in the NP domain. The first one is alignment of the gesture with an indefinite article, resulting in an interpretation of the gesture referent as being similar to the verbal referent it is temporally aligned with. Alternatively, the gesture can be aligned with a name or a definite article. In this case, this results in an interpretation where the gesture is identical to the verbal referent. Finally, aligning a gesture with an NP (i.e., a noun) results in an interpretation where the gesture referent exemplifies the verbal referent the gesture co-occurs with. Thus, the former two alignment patterns result in object-related interpretations, whereas the latter one results in a concept-related interpretation of the gesture in the sense of Fricke (2012). The different interpretation patterns and their corresponding interpretations have been experimentally validated by Ebert et al. (2022).

Ebert et al. (2020) adapt the formal system proposed by AnderBois et al. (2015), a unidimensional, dynamic system originally designed to capture relations between the two dimensions of meaning for appositives. However, since they argue that co-speech gestures behave akin to appositives, this system is particularly suitable to capture their meaning contribution adequately. To keep track of the at-issue/not-at-issue divide, they introduce the propositional variables p and $p*$ into the system. The variable p stands for an at-issue proposition, whereas $p*$ stands for a not-at-issue proposition.

In their formal system, variables, such as x, y, and z, stand for individual concepts and are thus of type $\langle s, e \rangle$. Direct reference to a gesture referent g is established by means of a rigid designator I_g, meaning that this rigid designator returns the same value (= the gesture referent) for all possible worlds. Finally, the coverbal performance of a gesture is formalized as in (6).

(6) $[z] \wedge z = I_g$

In (6), first an individual concept variable is introduced by means of the dynamic quantifier $[\cdot]$. Then, this variable is assigned the value of the rigid designator I_g.

Having introduced all the components essential to the system of Ebert et al. (2020), we can now turn to their formal proposal of a gesture that is aligned with an NP. Recall that the hypothesized interpretation is a concept-related one, which Ebert et al. (2022) dubbed an *exemplification reading*. The formalization of the sentence with an NP-aligned gesture in (7a) is as stated in (7b).

(7) a. Cornelia brought a [bottle].
BIG
b. $[z] \wedge z = I_g \wedge [x] \wedge \text{bottle}_p(x) \wedge \text{bottle}_{p*}(z) \wedge \text{bring}_p(x, \text{cornelia})$

In (7b), first an individual concept z is introduced that is assigned the value of the rigid designator I_g. This individual concept has the not-at-issue property of being a bottle. In other words, the gesture referent is itself a bottle according to Ebert et al. (2020). Then, a second individual concept is introduced which has the at-issue property of being a bottle. Finally, uttering the verb *brought* triggers the two-place at-issue relation that Cornelia brought x, i.e., the bottle. Crucially, the concept-related interpretation of the gesture comes about by means of the not-at-issue predicate $bottle_{p*}(z)$, as this predicate only specifies that the gesture referent exemplifies prototypical aspects of a bottle. Therefore, (7a) denotes two propositions. First, it denotes the at-issue proposition that Cornelia brought a bottle. Second, it denotes the not-at-issue proposition contributed by the gesture and its alignment to the speech signal that the gesture referent is itself a bottle.

As stated above already, aligning the gesture to an indefinite NP, as in (8a), results in an object-related interpretation. Consequently, the gesture referent in Ebert et al.'s (2020) system is not only a bottle itself, but has to resemble a real-world object (= a bottle), as well. This is modeled by introducing a similarity predicate adapted from Umbach and Gust (2014), the crucial difference being that the similarity predicate only takes two arguments, the entities which are similar to each other, and not three as suggested by Umbach and Gust (2014). Due to the similarity requirement, this reading is also called *comparison reading* of the gesture (cf. Ebert et al., 2022) and is formalized in (8b).

(8) a. Cornelia brought [a bottle].
BIG
b. $[z] \wedge z = I_g \wedge [x] \wedge \text{bottle}_p(x) \wedge \text{SIM}_{p*}(x, z) \wedge \text{bottle}_{p*}(z) \wedge \text{bring}_p(x, \text{cornelia})$

Comparing (8b) to (7b), one will notice that the only difference is the not-at-issue similarity predicate SIM, which is present in the former, but not the latter formula. This predicate establishes similarity between the bottle denoted by the verbal referent and the bottle denoted by the gesture referent. Thus, (8a) expresses two propositions: i) the at-issue proposition that Cornelia brought a bottle, which is the same as for (7a), and ii) the not-at-issue proposition that the gesture referent is itself a bottle and the bottle is similar to what is iconically depicted (i.e., it is large). Although we will not provide a formalism here, a gesture aligned with a proper name or a definite DP would result in a similar formalism as (8b), the only difference being that instead of a similarity relation, the verbal concept and the rigid designator denoted by the gesture would stand in a not-at-issue identity relation (cf. Ebert and Ebert, 2014; Ebert et al., 2020).

The hypothesis that different alignment patterns of gesture and speech give rise to different gesture interpretations has been experimentally validated by Ebert et al. (2022). To test for the hypothesis, they designed a rating study testing items as in (9).

(9) Jonas hat gestern ein seit Jahren unbewohntes Haus besichtigt.
'Yesterday, Jonas visited a house that has been unoccupied for years.'
a. Überraschenderweise sah er dort ein großes, sauberes Fenster.

'Surprisingly, he saw a big, clean window there.'

b. Es war also nicht überraschend, dass er dort kein großes, sauberes Fenster sah.
'It was hence no surprise that he saw no big, clean window there.'

(Ebert et al., 2022: 73)

Each item consisted of an introductory context sentence that remained constant across all conditions. There were two manipulations in the second sentence. First, it either contained an indefinite DP headed by the German indefinite article *ein* ('a') or an indefinite DP that contained the negative quantifier *kein* ('no') in the beginning. The second sentence always contained a gesture, in the case of (9), it was a rectangular gesture that provided information about the shape of the window. This gesture was either aligned with the whole DP, i.e., *(k)ein großes, sauberes Fenster* ('no/a big, clean window'), or with the NP *Fenster* ('window'). To make the different alignment patterns of gesture and speech easier to distinguish, two adjectives were inserted between the determiner and the noun. Since DP alignment of a gesture is hypothesized to give rise to a comparison reading, it should be illicit if the DP is headed by the negative quantifier *kein*. Their results corroborated this hypothesis: items in the condition where the gesture was DP aligned with *kein* were rated significantly worse than when the gesture was DP aligned with *ein*. Moreover, their analysis revealed no rating differences between the two conditions with NP alignment, which is also predicted by Ebert et al.'s (2020) proposal, since NP alignment only gives rise to an exemplification reading, which is available also for DPs headed by a negative quantifier.

Finally, Ebert et al. (2020) hypothesized and experimentally validated that demonstratives, such as the German demonstrative *so* ('such/like this') serve as what they call dimension shifters. This means that they are able to shift not-at-issue content to the at-issue dimension. It is defined as in (10). Thus, if a gesture is aligned with *so*, as in (10), it contributes its meaning in the at-issue dimension, as is evidenced by the fact that it can be directly denied:

(10) A: Ich habe [so eine Flasche Wasser] mitgebracht.
BIG
'I brought a bottle of water like this.
B: That's not true! The bottle is actually small.

A shortcoming of Ebert et al.'s (2020) supplemental account is that it focuses exclusively on speech-accompanying gestures in the nominal domain. An alternative proposal made by Schlenker (2018) treating co-speech gestures as a special kind of presupposition, namely co-suppositions, by contrast, strongly focuses on the VP domain, thereby making the two accounts difficult to compare in certain respects. However, we aim to change this by extending the supplemental account to the VP domain in this article. Before sketching our formal extension of Ebert et al. (2020), we will describe some new observations for gestures in the VP domain that have not yet been discussed in the formal semantic literature. We will then show how these findings can be implemented in our formal analysis.

3. Novel observations in the VP domain

As mentioned above already, gestures unsurprisingly also occur in the verbal domain. Consider the following example, where no square brackets are added, as we remain agnostic about the alignment patterns for now.

(11) Mary waved at Peter.
WAVING

The speaker of (11) has several options to perform the waving gesture that co-occurs with the verbal utterance. They could, for example, perform a gesture as the person in Figure 1 does. Since this gesture only incorporates those aspects which are prototypical for a waving event, i.e., only one hand is used to wave, etc., we will call this a *prototypical gesture*, which, as we will argue, is in large parts conventionalized. Alternatively, they could also perform a different type of gesture, such as a waving gesture where they use both hands. Since waving with both hands is a non-standard way of waving, we dub this type of gesture *non-prototypical gestures*. Crucially, these two types of gestures in the VP domain closely resemble the concept- and object-related interpretations of gestures in the NP domain observed by Fricke (2012). Before fleshing out the differences between these two types of gestures in more detail, we will elaborate on the notion of prototypical gestures.

Figure 1. A prototypical waving gesture (cf. Ortega and Özyürek, 2020).

In their study from 2020, Ortega and Özyürek investigated, among other things, whether there are systematic form-meaning mappings in iconic gestures. In other words, they investigated whether certain concepts are depicted systematically by speakers in (silent) gesture. To give an example, the concept *ball* could either be depicted by a gesture tracing the balls shape or alternatively by performing a throwing gesture indicating how the object is handled. Moreover, they investigated whether speakers tend to depict concepts within the same semantic domain by means of the same mode of representation (Müller, 2013; Hwang et al., 2017). They discuss four different modes of representation: representing, drawing, personification, and acting.

When using the representing strategy, the hand adopts the form of a referent. For example, when performing a gesture depicting someone walking down a flight of stairs, a speaker can use their fingers woggling diagonally downward in the gesture space. This strategy is often used when producing so-called observer viewpoint gestures, gestures which depict an event as if observed from a distance (McNeill, 1992). The drawing mode, by contrast, traces the form of an object. In other words, the form is painted in the gesture space. An example would be to trace the round shape of a ball. In personification, "the body serves as a map for a comparable non-human body" (Hwang et al., 2017: 574). A speaker can make use of this strategy when

depicting a bird by using their arms imitating a bird's wings. Finally, when using the acting strategy the body of the speaker represents itself, thereby depicting the handling of objects or intransitive actions, as, for example, the waving gesture in Figure 1. This strategy is typically employed when performing character viewpoint gestures. This type of gesture depicts an event from an internal first-person perspective (McNeill, 1992). Interestingly, there is a cross-cultural tendency to depict actions and objects by means of an acting strategy (Padden et al., 2013, 2015). We will therefore continue our discussion based on the waving gesture, as this gesture is itself an instance of the acting mode of representation.

To target their research questions, Ortega and Özyürek (2020) conducted two experimental studies. In the first experiment, a gesture generation task, 20 native speakers of Dutch were asked to produce silent gestures for different verbal concepts ($n = 272$). They were distinguishable into five different semantic domains: i) actions with objects, such as *to smoke*, actions without objects, for example, *to cry*, manipulable objects (*towel*), non-manipulable objects (*building*), and, finally, animate entities, such as *bear*. Each concept was presented on a computer screen for 4,000 ms. Crucially, in order to get intuitive results, participants had to come up with a silent gesture for each concept during this period of time. If they did not produce a gesture during this time frame, the study proceeded to the next trial.

In line with a norming system (Bressem, 2013; Ladewig and Bressem, 2013), the recordings of the silent gestures were classified with respect to four features: hand shape, orientation, movement, and placement. This system is based on a system used in sign language linguistics developed to classify lexical signs with respect to their phonological features, that is, hand shape, location, movement, and orientation (Brentari, 1999) and is argued to represent a gesture's most prominent features, with hand shape being the most prominent feature among the four.

Based on this annotated data set, Ortega and Özyürek (2020) wanted to determine the degree of systematicity among the produced silent gestures for each verbal concept. To achieve this, they took the data from all participants and compared for each concept the four features of each gesture for a given concept. A gesture was considered to be systematic of a concept, if at least three of the annotated features were the same in 50% of the data points per concept. Ortega and Özyürek (2020) used this set of systematic gestures for further analysis.

Out of the 272 concepts shown to participants, they found systematic gestures for 109 concepts, with an overall mean of 17.5 participants producing the same gesture for these concepts. Moreover, Ortega and Özyürek (2020) report that the acting strategy was predominantly used (70.64%), in line with previous research. An example is given in Figure 2. As can be easily

Figure 2. Different participants producing a gesture for the concept *telephone* (Ortega and Özyürek, 2020: 58).

observed, all four participants produce an identical acting gesture for the concept *relephone*.

In their second experiment, Ortega and Özyürek (2020) tackled the research question of how transparent the systematic gestures were they found in their first experiment, i.e., how well the gesture represents a verbal concept. To address this, they made video recordings of the 109 systematic gestures obtained from the first experiment and let participants rate their transparency on a 7-point scale. The results show that representing and acting had the highest transparency ratings (4.96 and 4.88, respectively), followed by personification (4.73) and drawing (3.43). Additionally, they found differences in transparency ratings between the aforementioned semantic categories within a given mode of representation, which we cannot discuss in detail here, as this would exceed the scope of the present article.

The question important to us is what the findings of Ortega and Özyürek's (2020) studies suggest concerning our discussion about prototypical and non-prototypical gesture interpretations in the VP domain. At this point, we would like to claim that our notion of a prototypical gesture interpretation shares important features with the notion of systematic gestures coined by Ortega and Özyürek (2020). The only important difference between the two terms is that we use our term in a more narrow sense than Ortega and Özyürek (2020) use theirs: while systematic gestures can occur in any part of the sentence, we limit the notion of prototypical gestures to the VP domain. In other words, a prototypical gesture is a systematic gesture occurring in the VP domain.

Turning now to interpretational differences between the two gesture types, consider again example (11), repeated in (12) for convenience.

(12) Mary waved at Peter.
WAVING

Consider now the two waving events sketched in Figure 3. When ignoring the gesture, the sentence in (12) can in principle be used to describe either event. Now note that when a single-handed waving gesture is produced, intuitively, the sentence can still adequately describe either event in Figure 3. This is due to the single-handed waving gesture being a gesture prototypical for a waving event. However, when a waving gesture is produced with both hands and co-occurs with (12), it can only refer to the scenario in Figure 3b, but no longer to the scenario in Figure 3a. This is due to the gesture being a non-prototypical gesture. These observations have several important implications for the semantic account that we develop in the next section.

(a) Single-handed waving event. (b) Waving event with both hands.

Figure 3. Two different waving scenarios for a waving event.

First, the observation that when producing a single-handed waving gesture, one can refer to both scenarios in Figure 3 suggests that prototypical gestures are in large parts not interpreted iconically, meaning that a prototypical gesture only describes prototypical aspects of the event denoted by the gesture referent and does not need to resemble an event that happened in the real world. Otherwise, one could not use the single-handed gesture when describing the event in Figure 3b. Conversely, the inability to describe the event in Figure 3a when a waving gesture produced with both hands co-occurs with (12) suggests that non-prototypical gestures must resemble an event that happened in the real world and are consequently interpreted iconically. We will follow Ebert et al. (2020) in positing that the two different interpretations of a gesture in the VP domain come about via different gesture-speech alignment patterns.

4. Formal proposal

To account for the differences observed between prototypical and non-prototypical gestures in Section 3, we first need to extend Ebert et al.'s (2020) supplemental account to the VP domain. A neat way to achieve this is by incorporating event semantics into their account. This has in parts already been done by Ebert and Hinterwimmer (2022), although their main aim was to account for quoted gestures which they view as demonstrations in the sense of Clark and Gerrig (1990), which have been argued to be special kinds of events in Davidson (2015).

They build on Davidson's (2015) approach to quotation, which follows suggestions made in Clark and Gerrig (1990) that quotation involves a demonstration. Davidson incorporates this idea into her formal account by introducing demonstrations, objects of type *d*, into the semantic ontology. Besides data from role shift phenomena in sign languages, she is mainly concerned with data from spoken language *be like* constructions in her paper. Examples are given in (13) and (14), respectively. She captures these observations in a neo-Davidsonian event semantics.

(13) Bob was like ["This isn't fair."]
WHINEY VOICE

(14) Bob was like [ ].[3]
GOBBLING

(Davidson, 2015: 489)

The whiney voice in (13) and the gobbling gesture in (14) can be directly denied, as attested by the felicity of the utterances in (15) (cf. Davidson, 2015: 489) as responses to the examples in (13) and (14), respectively.

(15) a. No, I think he wasn't being whiney when he said that.
b. No, he didn't gobble it down in such a way.

Interestingly, as Ebert and Hinterwimmer (2022) note, when transforming (13) to a direct discourse utterance (cf. (16)), the direct denial of Bob's way of speaking becomes infelicitous. The same has been observed for gestures, even outside of quotation (cf. Section 2 of this paper).

(16) Bob said: ["This isn't fair."]
WHINEY VOICE

[3]The empty square brackets indicate that the gobbling gesture was used pro-speech, i.e., speech-replacing.

In other words, the manner of speaking as well as gestures seem to be at-issue in *be like* constructions, but not-at-issue in direct discourse utterances. Consequently, quotative *like* is reminiscent of demonstratives, which have been shown to be dimension shifters by Ebert et al. (2020), meaning that they shift not-at-issue content to the at-issue dimension (cf. Section 2 of the present article). Thus, Ebert and Hinterwimmer (2022) conclude that the manner of speaking and also gestures, i.e., demonstrations, contribute not-at-issue meaning by default, evidenced by their behavior in direct discourse. When embedded under *be like*, they become at-issue, which Ebert and Hinterwimmer (2022) attribute to quotative *like* functioning as a dimension shifter (cf. also Streeck, 2002).

Since Davidson's (2015) account remains agnostic about the at-issue/not-at-issue divide, but the supplemental account of gesture semantics (Ebert et al., 2020) makes very straightforward predictions, Ebert and Hinterwimmer (2022) propose to combine the two. Crucially, they formally integrate a neo-Davidsonian event semantics account into the supplemental account proposed by Ebert et al. (2020), which is what we need to adequately capture the meaning contributions of gestures in the VP domain and the concomitant distinction between prototypical and non-prototypical gestures.

For their formal apparatus, Ebert and Hinterwimmer (2022) argue that gesture-speech interaction is what triggers the demonstration argument, or, in their case, the similarity predicate. The driving factor in their account is the interaction of gesture and speech. Concretely, they posit that expressions of type $\langle s, t\rangle$ (= (pseudo)-assertional sentences or clauses) make the semantic contribution that the reported event and the event demonstrated by the gesture must be similar. Note that they use the term gesture in a loose sense here, i.e., demonstrating that someone spoke with a whiney voice is also seen as a gestural contribution and consequently denotes a rigid designator I_g. This spares them from introducing a semantic type for demonstrations into their framework. They propose to formalize (13) and (16) as in (17a) and (17b), respectively (cf. Ebert and Hinterwimmer, 2022: 340).

(17) a. $[e] \wedge \text{agent}(e, \text{bob}) \wedge [z] \wedge z = I_g \wedge \text{SIM}_p(e, z)$

b. $[e] \wedge \text{agent}(e, \text{bob}) \wedge \text{say}(e) \wedge \text{form}_p(e) =$"*This isn't fair*"$\wedge [z] \wedge z = I_g \wedge \text{SIM}_{p*}(e, z)$

Setting aside the specification of the event being a saying event (triggered by the verb *said*) and the at-issue form contribution in (17b), the main difference between the two formalisms is that while the whiney voice demonstration denoted by I_g stands in a not-at-issue similarity relation to the reported event in (17b), this similarity relation is at-issue in (17a), i.e., the *be like* construction.

Transferring this implementation to our distinction between prototypical and non-prototypical gestures in the VP domain, we argue that prototypical gestures trigger a reading akin to an exemplification reading in the NP domain. Thus, the gesture only exemplifies some relevant components of the prototypical event it denotes. This results in the gesture not being interpreted as iconic, strictly speaking, as we predicted in Section 3. We dub this reading *event-kind reading*, as the gesture is not interpreted as referring to an event memorized by the speaker, but rather to an abstract event prototype. We follow Ebert et al. (2020) and assume that also in the VP domain, different alignment patterns of gesture and speech result in different interpretations of the gesture. In the verbal domain, aligning a gesture with the VP (cf. (18a)) results in an

event-kind interpretation. We formally model this as in (18b).

(18) a. Mary [waved] at Peter.
WAVING (SINGLE-HANDED)
b. [e] ∧ agent(e, mary) ∧ theme(e, peter) ∧ wave_{p}(e) ∧ [z] ∧ z = I_g ∧ wave_{p*}(z)

Thus, (19a) expresses the at-issue proposition that there exists an event such that Mary was the agent and Peter the theme of this event. Moreover, it expresses the not-at-issue proposition that the gesture aligned with the verb *waved* is a waving event itself and it displays relevant prototypical aspects of a waving event.

Turning to non-prototypical gestures next, we argue that they give rise to a reading that closely parallels a comparison reading in the NP domain, which we dub *event-token reading*. Recall that we have shown in Section 3 above that non-prototypical gestures make reference to an event that took place in the actual world, meaning that they must resemble this event in order to be felicitously used. We model this by means of the SIM predicate proposed by Ebert et al. (2020), stating that similarity has to hold between the event described by the sentence and the gesture referent. Crucially, an event-token interpretation comes about by means of aligning the gesture to the entire TP, that is, to the entire declarative sentence in (19a).

(19) a. [Mary waved at Peter].
WAVING (BOTH-HANDED)
b. [e] ∧ agent(e, mary) ∧ theme(e, peter) ∧ wave_{p}(e) ∧ [z] ∧ z = I_g ∧ wave_{p*}(z) ∧ SIM_{p*}(z, e)

Thus, we propose that (19a) expresses the same at-issue proposition as (18a) does, since it only contains the assertive content expressed by the verbal part of the utterance. Moreover, it expresses the not-at-issue proposition that the gesture referent is itself a waving event and that it is similar to the event introduced by the asserted part of the utterance.

Additionally, we propose that in the VP domain, the similarity that needs to hold between the gesture referent and the event is mediated by perspective dependence. In the remainder of this section, we will spell this intuition out formally. Before that, it is worth noting that a similar analysis has been proposed for so-called classifiers in sign languages by Schlenker and Lamberton (2024). They argue that the interpretation of classifiers in sign languages has a twofold nature: It contains a lexical component as well as a component that is interpreted iconically, i.e., the movement of the classifier and its position in the signing space. To model the partly iconic interpretation of a classifier, they make use of a projection function (see, e.g., Abusch, 2020). Schlenker and Lamberton (2024) propose the following truth conditions for the classifier in American Sign Language that represents an upright person:

(20) $[\![\text{person-cl}^{\pi}_{a}]\!]^{c, s, t, w}$ = 1 iff $\text{person'}_{t, w}(s(a)) = 1$ and proj(s(a), s(π), t, w) = person-cl
(Schlenker and Lamberton, 2024)

Thus, in their analysis, the person classifier is true in a world *w* at time *t* iff i) the locus *a* assigned by the assignment function *s* is a person at *t* in *w* and ii) this locus *a* projects to *person-cl* at time *t* in *w* from viewpoint π assigned by *s*.

Transferring the second part of this lexical entry to the case of a non-prototypical gesture in the VP domain, we propose the following truth conditions for SIM:

(21) $\text{SIM}_{p^*}(z, e) = 1$ iff $\forall w[p^*(w) \rightarrow \exists \pi[\text{proj}(\pi, e, w) = z]]$

We thus posit that similarity between the gesture referent z and the event e holds with respect to $p*$ iff for all worlds in which $p*$ is true and in which e therefore occurred (i.e., of which e is a part), there is a viewpoint π such that e projects to z in w from viewpoint π. This condition holds if seen from a certain perspective (from a position facing Mary while she waved at Peter, for instance), the waving gesture performed by Mary looks like the gesture performed by the speaker while uttering the sentence. Note that the projection function assumed here may be rather coarse grained, i.e., it need not map all the details of Mary's original waving gesture to the waving gesture performed by the speaker, but rather retain some noteworthy features, such as it being performed with both hands rather than one hand. Our proposal currently makes an interesting prediction regarding the alignment possibilities of prototypical and non-prototypical gestures, namely that non-prototypical gestures must be aligned with the TP, whereas prototypical gestures can either be aligned with the VP or the TP. The explanation behind this is as follows: Non-prototypical gestures are normally modified prototypical gestures. In the case of the waving gestures used throughout the article, using both hands to depict waving is clearly a modification of the prototypical single-handed waving, as the use of both hands to depict waving clearly adds further information in comparison to the prototypical, single-handed case. Aligning a prototypical gesture to the TP (i.e., single-handed waving) is, as mentioned above, intuitively unproblematic, as it can either be interpreted as an event kind or an event token. However, when aligning a non-prototypical gesture to the VP, we predict that this results in an event-kind interpretation of this gesture, which should result in infelicity. While this is an interesting prediction, this should be investigated experimentally, as it is unclear whether this prediction is empirically adequate.

5. Conclusion

In this paper, we have proposed an analysis of the different ways in which co-speech gestures depicting prototypical as opposed to co-speech gestures depicting non-prototypical ways of performing actions are interpreted. While the former remain felicitous when they accompany verbal descriptions of events that actually differ in relevant respects from the co-speech gesture, the latter are only felicitous when the reported event actually was similar in relevant respects to the co-speech gesture. Our analysis extends Ebert et al.'s (2020) distinction between concept- and object-related interpretations of gestures to the VP domain, arguing that prototypical gestures depicting events denote event types, while non-prototypical gestures depicting events denote event tokens. Therefore, while the latter do not really provide additional information about the event introduced by the respective verbal utterance, the latter provide such information and are therefore only felicitous if the depicted event is actually similar in relevant respects to the co-speech gesture. This similarity requirement is formally captured by the assumption that in all worlds in which the respective event e occurred there is viewpoint such that the event from that viewpoint looks in relevant respects like the gesture. In future work, we are planning to return to a prediction that the analysis proposed in this paper makes, namely that non-prototypical gestures must be aligned to the TP, while prototypical gestures can be aligned to the VP or the TP.

References

Abusch, D. (2020). Possible worlds semantics for pictures. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann, and T. E. Zimmermann (Eds.), *The Wiley Blackwell Companion to Semantics*. Hoboken, NJ: Wiley Blackwell.

AnderBois, S., A. Brasoveanu, and R. Henderson (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics 32*(1), 93–138.

Brentari, D. (1999). *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.

Bressem, J. (2013). A linguistic perspective on the notation of form features in gestures. In N. J. Enfield, C. Müller, E. Fricke, S. Ladewig, A. Cienki, D. McNeill, and S. Teßendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, pp. 1079–1098. Berlin, Germany & Boston, MA: De Gruyter Mouton.

Chierchia, G. and S. McConnell-Ginet (1990). *Meaning and Grammar*. Cambridge, MA: MIT Press.

Clark, H. H. and R. J. Gerrig (1990). Quotations as demonstrations. *Language 66*(4), 764–805.

Davidson, K. (2015). Quotation, demonstration, and iconicity. *Linguistics and Philosophy 38*(6), 477–520.

de Ruiter, J. P. (1998). *Gesture and Speech Production*. Ph. D. thesis, University of Nijmegen, Nijmegen, The Netherlands.

Ebert, C. (2024). Semantics of gesture. *Annual Review of Linguistics 10*(1), 169–189.

Ebert, C. and C. Ebert (2014). Gestures, demonstratives, and the attributive/referential distinction. Talk given at *Semantics and Philosophy in Europe 7*.

Ebert, C., C. Ebert, and R. Hörnig (2020). Demonstratives as dimension shifters. In *Proceedings of Sinn und Bedeutung (SuB) 24*, Osnabrück, Germany, pp. 161–178. University of Osnabrück.

Ebert, C., S. Evert, and K. Wilmes (2011). Focus marking via gestures. In I. Reich, E. Horch, and D. Pauly (Eds.), *Proceedings of Sinn und Bedeutung 15*, pp. 193–208. Saarbrücken, Germany: Saarland University.

Ebert, C. and S. Hinterwimmer (2022). Free indirect discourse meets character viewpoint gestures. In S. Featherston, R. Hörnig, A. Konietzko, and S. von Wietersheim (Eds.), *Proceedings of Linguistic Evidence 2020: Linguistic Theory Enriched by Experimental Data*, Tübingen, Germany, pp. 333–349. University of Tübingen.

Ebert, C., G. Pirillo, and S. Walter (2022). The role of gesture-speech alignment for gesture interpretation. In S. Featherston, R. Hörnig, A. Konietzko, and S. von Wietersheim (Eds.), *Proceedings of Linguistic Evidence 2020*, Tübingen, Germany, pp. 65–77. University of Tübingen.

Esipova, M. (2019). *Composition and Projection in Speech and Gesture*. Ph. D. thesis, New York University, New York, NY.

Farkas, D. F. and K. B. Bruce (2010). On reacting to assertions and polar questions. *Journal of Semantics 27*(1), 81–118.

Fricke, E. (2012). *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin, Germany & New York, NY: Walter de Gruyter.

Hwang, S.-O., N. Tomita, H. Morgan, R. Eegin, D. İlkbaşaran, S. Seegers, R. Lepic, and C. Padden (2017). Of the body and the hands: Patterned iconicity for semantic categories. *Language and Cognition 9*(4), 573—602.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication*, pp. 207–227. The Hague, The Netherlands: Mouton.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.

Ladewig, S. H. and J. Bressem (2013). A linguistic perspective on the notation of gesture phases. In N. J. Enfield, C. Müller, E. Fricke, S. Ladewig, A. Cienki, D. McNeill, and S. Teßendorf (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*, pp. 1060–1079. Berlin, Germany & Boston, MA: De Gruyter Mouton.

Langendoen, D. T. and H. Savin (1971). The projection problem for presuppositions. In C. J. Fillmore and D. T. Langendoen (Eds.), *Studies in linguistic semantics*, pp. 54–60. New York, NY: Irvington.

Loehr, D. P. (2004). *Gesture and Intonation*. Ph. D. thesis, Georgetown University, Washington, DC.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.

Müller, C. (2013). Gestural modes of representation as techniques of depcition. In N. J. Enfield, C. Müller, E. Fricke, S. Ladewig, A. Cienki, D. McNeill, and S. Teßendorf (Eds.), *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*, pp. 1687–1701. Berlin, Germany & Boston, MA: De Gruyter Mouton.

Nunberg, G. (1993). Indexicality and deixis. *Linguistics and Philosophy 16*(1), 1–43.

Ortega, G. and A. Özyürek (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: A normed database of silent gesture. *Behavior Research Methods 52*(1), 51–67.

Padden, C. A., S.-O. Hwang, R. Lepic, and S. Seegers (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in Cognitive Science 7*(1), 81–94.

Padden, C. A., I. Meir, S.-O. Hwang, R. Lepic, S. Seegers, and T. Sampson (2013). Patterned iconicity in sign language lexicons. *Gesture 13*(3), 287–308.

Potts, C. (2005). *The Logic of Conventional Implicatures*. Oxford, UK: Oxford University Press.

Schlenker, P. (2018). Gesture projection and cosuppositions. *Linguistics and Philosophy 41*(3), 295–365.

Schlenker, P. and J. Lamberton (2024). Iconological semantics. *Linguistics and Philosophy 47*, 779–838.

Shanon, B. (1976). On the two kinds of presuppositions in natural language. *Foundations of Language 14*(2), 247–249.

Streeck, J. (2002). Grammars, words, and embodied meanings: On the uses and evolution of *so* and *like*. *Journal of Communication 52*(3), 581–596.

Tonhauser, J. (2012). Diagnosing (not-)at-issue content. In E. Bogal-Allbritten (Ed.), *Proceedings of Semantics of Under-represented Languages of the Americas (SULA) 6*, pp. 239–254.

Umbach, C. and H. Gust (2014). Similarity demonstratives. *Lingua 149*, 74–93.

von Fintel, K. (2004). Would you believe it? The king of France is back! In A. Bezuidenhout and M. Reimer (Eds.), *Descriptions and Beyond*, pp. 315–347. Oxford, UK: Oxford University Press.