frontiers in Communication

ORIGINAL RESEARCH
published: 15 June 2021
doi: 10.3389/fcomm.2021.625757

On the Interaction of Gestural and Linguistic Perspective Taking

Stefan Hinterwimmer[1]*, *Umesh Patil*[2] *and Cornelia Ebert*[3]

[1]Department of German Studies, Bergische Universität Wuppertal, Wuppertal, Germany, [2]Department of German Language and Literature I, Universität zu Köln, Cologne, Germany, [3]Department of Linguistics, Goethe-Universität Frankfurt, Frankfurt, Germany

In this paper, we investigate the question of whether and how perspective taking at the linguistic level interacts with perspective taking at the level of co-speech gestures. In an experimental rating study, we compared test items clearly expressing the perspective of an individual participating in the event described by the sentence with test items which clearly express the speaker's or narrator's perspective. Each test item was videotaped in two different versions: In one version, the speaker performed a co-speech gesture in which she enacted the event described by the sentence from a participant's point of view (i.e. with a character viewpoint gesture). In the other version, she performed a co-speech gesture depicting the event described by the sentence as if it was observed from a distance (i.e. with an observer viewpoint gesture). Both versions of each test item were shown to participants who then had to decide which of the two versions they find more natural. Based on the experimental results we argue that there is no general need for perspective taking on the linguistic level to be aligned with perspective taking on the gestural level. Rather, there is clear preference for the more informative gesture.

Keywords: perspective taking, free indirect discourse, viewpoint, co-speech gestures, observer viewpoint gestures, character viewpoint gestures, anti-logophoricity

OPEN ACCESS

Edited by:
Jorrig Vogels,
University of Groningen, Netherlands

Reviewed by:
Marisa Casillas,
Max Planck Institute for Psycholinguistics, Netherlands
Gisela Redeker,
University of Groningen, Netherlands
Natalie Dowling,
University of Chicago, United States,
in collaboration with reviewer MC

***Correspondence:**
Stefan Hinterwimmer
hinterwimmer@uni-wuppertal.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *03 November 2020*
Accepted: *24 May 2021*
Published: *15 June 2021*

Citation:
Hinterwimmer S, Patil U and Ebert C (2021) On the Interaction of Gestural and Linguistic Perspective Taking. Front. Commun. 6:625757. doi: 10.3389/fcomm.2021.625757

INTRODUCTION

Perspective taking is an integral part of the information conveyed by sentences that are contained in narrative texts. One and the same event can be described from a detached observer's perspective or as if it was perceived by a participant. Free Indirect Discourse (FID) is a particularly clear way for a sentence to express the perspective of a protagonist in a narrative text. In FID, all perspective-dependent expressions (i.e. deictic expressions such as tomorrow and here, evaluative expressions, interjections etc.) are interpreted from the perspective of some contextually salient protagonist, while pronouns and tenses are interpreted from the (possibly entirely abstract) narrator's perspective (Rauh 1978; Banfield 1982; Doron 1991; Schlenker 2004; Eckardt 2014; Maier 2015). Additionally, the proposition denoted by a sentence in FID is interpreted as a thought or utterance that the respective protagonist has or makes at the reference time of the ongoing story (Eckardt 2014). The second sentence in (1), for instance, which is a clear instance of FID, is thus interpreted as a thought that Masha has at the time of her staring at Wilfred–the same thought as the one rendered as direct discourse in (1b).

(1) a. Masha stared at Wilfred in disbelief. Had that idiot really invited her to his birthday party tomorrow evening?

b. Masha stared at Wilfred in disbelief. She thought: 'Has that idiot really invited me to his birthday party tomorrow evening?'

At the same time, it is well-known that perspective taking can also be expressed at the level of co-speech gestures, i.e. gestures that speakers produce while uttering sentences. In particular, there are two kinds of iconic gestures that are often used by speakers when they describe scenes or events to their interlocuters and that clearly reveal a perspective: character viewpoint gestures (CVG), on the one hand, and observer viewpoint gestures (OVG), on the other (McNeill, 1992; Parrill, 2010, Parrill, 2012; Stec, 2012, Stec, 2016). When performing the former, the speaker impersonates an individual participating in the event described by the sentence and enacts the event from that person's point of view by using her entire body in combination with facial expressions. When performing the latter, in contrast, the speaker depicts the event described by the sentence as if it was observed from a distance, usually by using the hands exclusively which then represent a participant, with the hand's trajectory representing that participant's path, for instance.

Based on Parrill (2010) assumption that gestural and linguistic viewpoint have a common conceptual source, we investigate the question in this paper of whether and how perspective taking at the linguistic level interacts with perspective taking at the level of co-speech gestures. In an experimental study conducted in German we compared test items clearly expressing the perspective of an individual participating in the events described with test items which clearly express the speaker's or narrator's perspective. Test items of the former kind were always construed in such a way that the opening sentence describes the feelings, intentions or thoughts of a protagonist in a particular situation which the second sentence specifies in more detail, while the final sentence renders a thought that the respective protagonist has in that situation in the form of FID. Test items of the latter kind, in contrast, always contained an opening sentence in the form of a general statement in present tense expressing an evaluation of or opinion about the individual participating in the event described by the following sentences. Additionally, that individual was referred to by a demonstrative pronoun. As shown by (Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017), German demonstrative pronouns are anti-logophoric pronouns, i.e. they cannot refer to an individual whose perspective is expressed by the sentence containing them.

Each test item was videotaped in two different versions: In one version, the speaker performed a CVG while uttering the respective final sentence. In the second version, the speaker performed an OVG. Both versions of each test item were shown to participants who then had to decide which of the two versions they find more natural. If Parrill (2010) is right in assuming that gestural and linguistic viewpoint have a common conceptual source, it is to be expected that those versions of the test items are preferred where the linguistically expressed perspective aligns with the perspective expressed on the gestural level. This expectation is based on the assumption that combinations of speech and gesture convey a complex multimodal message including a perspective that is planned by a central cognitive process and then dispatched into disparate channels (see De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; Kendon, 2004). Crucially, while the information conveyed via the two channels may be either redundant (thus avoiding misunderstanding) or complementary, the default should be for it to be coherent (see also Cassell et al., 1999, but see Goldin-Meadow, 1999 for arguments that mismatches between speech and gestures are sometimes productive and useful). Concerning perspective, this means that at least in the absence of intervening factors there should be a preference for perspective alignment. Consequently, the prediction is that participants choose the CVG-version of test items instantiating FID and thus expressing the participant's perspective more often than the OVG-version. For test items expressing the speaker's or narrator's perspective, in contrast, it should be the other way around, i.e. participants should choose the OVG-version more often than the CVG-version.

An exception to this general reasoning would be the case of multiple perspectives. A speaker might plan to express more than one perspective at the same time. In this vein, Parrill (2009) investigates dual viewpoint gestures (first noted by McNeill, 1992, see also Cassell et al., 1999 for discussion), i.e. gestures that simultaneously express more than one viewpoint. If gestures can express more than one viewpoint at the same time, it should also be possible to express more than one viewpoint in gesture and speech. For example, one could express a certain viewpoint A in speech and a certain viewpoint B in gesture, or, a viewpoint A in speech and a dual viewpoint expressing A and B in gesture. This possibility might have confounded our results and we will come back to it below.

As we will see, the predictions that FID couples with CVGs and linguistic narrator's perspective with OVGs are only partially confirmed by the experimental results: While CVGs were generally preferred in combination with both kinds of test items, this tendency was stronger in test items expressing the respective participant's perspective than in test items expressing the speaker's or narrator's perspective. We take this as evidence that there is no general need for perspective taking on the linguistic level to be aligned with perspective taking on the gestural level, contrary to our expectations. Nevertheless, there seems to be a general preference for more informative gestures, with CVGs being more informative than OVGs insofar as they are less schematic and more detailed. Since a speaker might wish to express dual viewpoints, she might choose to express one viewpoint in speech and another one, possibly one that is more informative, with gesture.

The paper is structured as follows. In *Free Indirect Discourse* and *Demonstrative pronouns as anti-logophoric pronouns*. we provide some theoretical background on FID and the anti-logophoricity of German demonstrative pronouns, respectively, and in *Observer viewpoint and character viewpoint gestures* on CVGs and OVGs. The experimental study is presented and discussed in *The experimental study*. *Conclusion* concludes the paper.

THEORETICAL BACKGROUND

Free Indirect Discourse

As already said in the introduction, FID is a particularly clear form of linguistically encoded perspective taking that is largely confined to narrative texts (but see Redeker, 1996 for evidence that FID is often found in journalistic texts as well and Stokke, 2020 for discussion of the use of FID in non-fictional texts such as biographies of historical figures; see Brinton, 1980, Banfield, 1982, Stokke, 2013, Hinterwimmer, 2017 and Abrusán, 2020 for another form of linguistically encoded perspective taking dubbed free perception report, viewpoint shifting, or protagonist projection). A sentence in FID conveys a character's thoughts or utterances without there being any overt indication of a shift from the (potentially entirely abstract) narrator's external perspective to the internal perspective of the respective character. An overt indication of such a shift would be the presence of a propositional attitude verb, where the external argument slot is occupied by a phrase referring to that character in combination with quotation marks, as in direct discourse (DD). What FID shares with DD is the interpretation of local and temporal deictic expressions with respect to the context of the individual whose thoughts or utterances are rendered. At the same time, pronouns and tenses are interpreted with respect to the narrator's context, as in indirect discourse. Consider again (1a) from above, repeated here as (2a): The second sentence is interpreted as a question that Masha asks herself while she is staring at Wilfred in disbelief. Crucially, the deictic temporal adverb tomorrow, which is usually interpreted with respect to the context of utterance (Kaplan 1989), is interpreted in the same way in (2a) as it is interpreted in (2b), namely as referring to the day following the day on which she has the thought reported by the respective sentence. In contrast to (2b), however, where that thought is rendered as DD, Masha is referred to by the third person pronoun *her* instead of the first person pronoun *me*, and the auxiliary verb *have* is marked for past instead of present tense.

(2) a. Masha stared at Wilfred in disbelief. Had that idiot really invited her to his birthday party tomorrow evening?
b. Masha stared at Wilfred in disbelief. She thought: "Has that idiot really invited me to his birthday party tomorrow evening?"

Deictic expressions thus do not behave uniformly in FID: While the vast majority of them is interpreted with respect to the context set up by the preceding sentence, pronouns and tenses are always interpreted with respect to the narrator's context. Concerning evaluative expressions such as *that idiot* in (2a), interjections such as *wow*, *ouch* and *oops* or exclamatives, they are always interpreted with respect to the respective character's perspective–in (2a), for instance, it is Masha who considers Wilfred an idiot, not the narrator. Likewise, the exclamative in (3) expresses Tom's delight and surprise in virtue of the extent to which he is (or rather believes himself to be) smart, not the narrator's.

(3) Tom leaned back in his chair, smiling at the man he had just cheated out of 5,000 dollars. Wow, how smart he was!

Concerning the question of how the distinctive properties of FID just outlined are to be captured, there are two lines of analysis that have been proposed in the formal semantics literature: double context analyses (Schlenker 2004, Sharvit 2008, Eckardt 2014; see Rauh 1978, Banfield 1982 and Doron 1991 for earlier implementations of similar ideas in different frameworks) and the mixed quotation approach. The basic idea behind double context analyses is that sentences in FID are interpreted not just with respect to a context of utterance (Kaplan 1989), but with respect to two different contexts. Concerning the technical details, the following exposition is based on Eckardt (2014). Eckardt assumes that while ordinary utterances in everyday communication are just interpreted with respect to a context of utterance *C*, narrative texts allow the addition of a second context *c*, with *C* being the context of the (potentially entirely abstract) narrator and *c* being the context of some character that is prominent at that point in the discourse (see Hinterwimmer 2019 and Hinterwimmer and Meuser 2019 for detailed discussion of the conditions under which characters are prominent enough to serve as potential anchors for FID). This second context *c* is implicitly introduced by the preceding discourse and it consists of the character functioning as the author (i.e. the speaker or thinker) of *c* and the spatial and temporal location of that character at the reference time of the ongoing story.

Whenever only *C* is present, all context-sensitive expressions are interpreted with respect to *C*. Crucially, however, all context-sensitive expressions with the exception of pronouns and tenses have to be interpreted with respect to *c* whenever *c* is available, while pronouns and tenses always have to be interpreted with respect to *C*. Additionally, whenever a sentence is interpreted both with respect to *C* and *c*, the proposition *p* (or set of propositions *Q* in cases such as (2a), where the sentence in FID is a polar question) it denotes is not directly added to the set of propositions that characterize the fictional story worlds. Rather, the proposition that the author of *c* believes *p* (or asks herself *Q*) is added, thus ensuring that sentences in FID are interpreted as the respective character's thoughts rather than assertions by the narrator.

Concerning the second sentence in (2a), for instance, the deictic temporal adverb tomorrow is interpreted with respect to the temporal parameter of *c*, i.e. as referring to the day following the day on which Masha stared at Wilfred in disbelief, and the negative evaluation of Wilfred as an idiot is attributed to the author of *c*, i.e. Masha. The past tense marking of the auxiliary verb *have* and the third person features on the pronouns her and him, in contrast, are interpreted with respect to the narrator's context *C*, thus requiring temporal location before the time of *C* rather than *c* and distinctness from the author (and addressee) of *C* rather than *c* (which is why *her* can be interpreted as referring to Masha), respectively. Finally, the sentence is interpreted as a question that Mary is asking herself at the time of *c*, i.e. at the time of her staring at Wilfred in disbelief.

A fundamentally different analysis of FID is proposed by Maier (2015, 2017: see also Dirscherl and Pafel, 2015). On this approach, FID is a special, highly conventionalized form of mixed quotation: Sentences in FID are quotes of thoughts or utterances, with tenses and pronouns being unquoted. In contrast to more familiar forms of mixed quotation, as they are found in newspaper articles, the quoted parts are not typographically marked as such (via quotation marks or italics, for instance) and there is no introductory clause such as *x said/thought* signaling that the following clause is to be interpreted as the partial quote of a thought or utterance of *x*. On the mixed quotation approach (2a) corresponds to the (simplified) schematic representation in (4a) and is interpreted as paraphrased (in simplified form) in (4b), the idea being that speaking and thinking events have both a form and a content and can be decomposed into subevents corresponding to parts of the respective thought or utterance.

(4) a. Mary stared at Wilfred in disbelief. Had 'that idiot really invited' her 'to his birthday party tomorrow evening'? b. There is an event *e* of Mary staring in disbelief that is located before the time of *C* and there is an event e_1 of Masha thinking that is located at the time of *e* and there are subevents e_2 and e_3 of e_1, and the form of e_1 is the form of e_2 concatenated with *that idiot really invited* concatenated with the form of e_3 concatenated with *to his birthday party tomorrow evening*, and the content of e_2 is the denotation of had and the content of e_3 is the denotation of her.

Concerning the question of why pronouns and tenses are systematically unquoted in FID (Maier, 2015, Maier, 2017), assumes this to be the result of a pragmatic tendency that can be observed in other forms of mixed quotation as well and has become fully conventionalized in the case of FID. Note that in spite of the profound differences in technical implementation, the resulting interpretations are rather similar to those assumed by the double context analyses: A speaking or thinking event has to be accommodated by the reader and the content of this event is the semantic object denoted by the respective sentence in FID. The quoted context-sensitive expressions are ultimately interpreted with respect to the context of the character whose thought or utterance is being quoted, while the unquoted ones receive their standard interpretation with respect to the narrator's context. Nevertheless, there is an argument in favor of the mixed quotation approach: As pointed out by Maier (2015), there are cases of FID where a character's thoughts or utterances are rendered in the non-standard dialect spoken by that character, while the surrounding text is written in standard language. While such cases are easily accounted for in the mixed quotation approach, it is hard to see how they could be captured by double context analyses.

Having discussed FID as a phenomenon where the perspective of some character becomes highly prominent (without the narrator's perspective disappearing completely, as evidenced by the interpretation of pronouns and tenses), we will discuss the use of German demonstrative pronouns in the following section and argue that these pronouns indicate that the speaker's or narrator's perspective is highly prominent.

Demonstrative Pronouns as Anti-logophoric Pronouns

Similarly to languages such as Dutch, Finnish and Catalan (see, e.g. Kaiser and Trueswell, 2008; Kaiser, 2010, Kaiser, 2011a, Kaiser, 2011b, Kaiser, 2013; Mayol and Clark, 2010), German does not only have personal pronouns (henceforth: PPros), but also demonstrative pronouns. The latter come in two varieties: the *der/die/das* series and the *dieser/diese/dieses* series. Since *diese* pronouns are largely confined to the formal register (see Patil et al., 2020 for recent discussion), we will set them aside for the purposes of this paper and concentrate on the contrast between PPros and demonstrative pronouns of the *der/die/das* variety, which we will henceforth refer to as DPros.

In the past, research on the contrast between DPros and PPros has mostly focused on cases like (5) (adapted from Bosch et al., 2007), where two potential antecedents with congruent gender features have been introduced in the preceding linguistic context and where there is genuine ambiguity in the resolution options of the pronouns. In such cases, there is a strong tendency for DPros to pick up the less prominent and for PPros to pick up the more prominent antecedent. Prominence has been defined in terms of (grammatical) subjecthood (Bosch et al., 2007; Hinterwimmer and Brocher 2018), topicality (Bosch and Umbach 2006; Hinterwimmer 2015), and (proto-)agentivity (Schumacher et al., 2016, Schumacher et al., 2017).

(5) Paul_i wollte mit Peter_j laufen gehen. Aber $\{\text{er}_{i,j}/\text{der}_j\}$ war leider erkältet.

Paul_i wanted to go running with Peter_j. But he $\{\text{PPro}_{i,j}/\text{DPro}_j\}$ had a cold unfortunately *(adapted from Bosch et al., 2007).*

In (5), Paul is more prominent than Peter since the proper name referring to him is the grammatical subject of the preceding sentence and the agent of the verb contained in that sentence, while the proper name referring to Paul is (contained in) the prepositional object and the theme. Additionally, Peter is also the aboutness topic (Reinhart, 1981) of that sentence by default, because the proper name referring to him is the subject of the sentence and because it occupies the leftmost position of the sentence. Consequently, there is a very strong tendency for the DPro *der* in (5) to pick up Peter, while the PPro *er* can be understood as picking up either Paul or Peter, with a (comparatively weak) preference for Paul.

As observed by Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017; see Hinterwimmer et al., 2020 and Hinterwimmer to appear for additional empirical evidence), however, there are cases like (6a) where referents that are maximally prominent in terms of subjecthood, topicality and (proto-)agentivity can nevertheless easily be picked up by DPros. At the same time, the contrast with (6b), where this is impossible or at least leads to rather strong markedness, shows that it is not the case that DPros are only prohibited from picking up maximally prominent referents in cases of potential ambiguity, i.e. whenever there is more than one potential antecedent with matching gender features available.

(6) Peter$_i$ seufzte, als er die Tür öffnete, und sah, dass die Wohnung mal wieder in einem fürchterlichen Zustand war.
Peter$_i$ sighed when he opened the door and saw that the flat was in a terrible state again.
a. Der $_i$ kann sich einfach nicht gegen seinen Mitbewohner durchsetzen.
He(DPro$_i$) is simply unable to stand his ground against his flatmate.
b. Verdammt, der$_{*i}$/er$_i$ hatte doch gestern erst aufgeräumt.
Damn, he(DPro$_{*i}$)/he$_i$ had only tidied up yesterday, after all. (Hinterwimmer et al., 2020: 114, ex. (8))

Peter is not only the subject and the agent of the matrix as well as the temporal adjunct clause in the opening sentence in (6), but its referent is presumably also the aboutness topic (Reinhart 1981) of the sentence, i.e. the proposition denoted by that sentence is understood as being about Peter rather than the door or the flat. If DPros avoid maximally prominent referents, Peter should thus be unavailable as an antecedent not only for the DPro in (6b), but also for the one in (6a). At the same time, if DPros were only prohibited from picking up maximally prominent referents in cases of potential ambiguity, Peter should not only be available as an antecedent for the DPro in (6a), but also in (6b). So, what distinguishes the continuation of (6) in (6a) from the one in (6b)?.

According to Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017, it is the fact that (6a) can only be interpreted as expressing the narrator's (or speaker's, if it is uttered in oral conversation) perspective, while (6b) is most plausibly interpreted as expressing Peter's perspective. In other words, while the narrator is the perspectival center with respect to the proposition denoted by (6a), Peter is the perspectival center with respect to the proposition denoted by (6b). In the case of (6a), it is the switch from past tense in the opening sentence, which is an instance of neutral narration, to present tense in the continuation, which breaks narrative continuity and in combination with the content establishes the narrator (or speaker) as the perspectival center. Consequently, (6a) is understood as a general statement about Peter's character by the narrator.

The continuation in (6b), in contrast, is most likely understood as a thought of Peter rendered in FID, i.e. the most plausible reading is one according to which Peter thought I only tidied up yesterday, after all when he saw the chaos in the kitchen. This is indicated by the content in combination with the presence of the deictic temporal adverb *gestern* ("yesterday"), the evaluative expression *verdammt* ("damn") and the modal particle *doch*. Concerning the deictic temporal adverb *gestern*, it is most plausibly interpreted as referring to the day preceding the day on which Peter came home in the evening, i.e. with respect to Peter's context, not with respect to the narrator's (or speaker's) context. Likewise, the evaluative expression *verdammt* ("damn") is more plausibly interpreted as expressing Peter's rather than the narrator's (or speaker's) frustration. Finally, the modal particle *doch*, which (very roughly) indicates that the proposition denoted by the clause containing it violates a previously held assumption, is more plausibly interpreted as violating Peter's rather than the narrator's (or speaker's) expectations.

From contrasts like the one between (6a) and (6b) in the context of (6), Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017 draw the conclusion that DPros avoid discourse referents functioning as perspectival centers as antecedents, where the term perspectival center is defined as in (7).

(7) A discourse referent α is the perspectival center with respect to a proposition *p* if *p* is the content of a mental state of the semantic value of α (i.e. g(*α*), where g is the assignment function).

Since the proposition denoted by (6b) (on its most plausible interpretation) is the content of a thought of Peter, Peter is the perspectival center with respect to that proposition and can accordingly not be picked up by a DPro, but only by a PPro (PPros being neutral in this respect, i.e. they can, but do not have to pick up discourse referents functioning as perspectival centers). Concerning the proposition denoted by (6a), in contrast, there is a strong tendency for it to be interpreted as the content of a thought of the narrator (or speaker). Consequently, the narrator (or speaker) is the perspectival center with respect to that proposition, and the DPro can accordingly be interpreted as picking up Peter, in spite of him being maximally prominent in terms of subjecthood (proto-)agentivity, and topicality. Because of their avoidance of perspectival centers, DPros are dubbed anti-logophoric pronouns in Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017 (see Charnavel and Mateu, 2015 and Yashima, 2015 for arguments that anti-logophoric pronouns exist in French, Spanish and Japanese, as well), i.e. the counterparts of pronouns existing in many West African and Asian languages that can only be used to pick up discourse referents functioning as perspectival centers and have been dubbed logophoric pronouns (Clements, 1975; Sells, 1987; Sundaresan, 2012; Nishigauchi, 2014; Pearson, 2015).

But what about cases like (5), where anti-logophoricity does not obviously play a role regarding the resolution options of DPros? Hinterwimmer and Bosch, 2016, Hinterwimmer and Bosch, 2017 argue that in the absence of the speaker or narrator functioning as perspectival center (i.e. in instances of neutral narration) the respective aboutness topic is the perspectival center by default, i.e. the proposition denoted by the respective sentence is interpreted as the content of a mental state of the topical referent, where that mental state need not be a conscious thought but can also be a state of perceiving. Evidence for this assumption is provided by the following observation (see Hinterwimmer et al., 2020 for additional empirical evidence): In the variant of (5) (repeated here as (8a)) given in (8b), where the second sentence is construed in such a way that it clearly expresses an evaluative comment by the speaker or narrator, the DPro can easily be understood as picking up Paul.

(8) a. Paul$_i$ wollte mit Peter$_j$ laufen gehen. Aber {er$_{i,j}$/der$_j$} war leider erkältet.
Paul$_i$ wanted to go running with Peter$_j$. But he {PPro$_{i,j}$/DPro$_j$} had a cold unfortunately.
b. Paul$_i$ wollte mit Peter$_j$ laufen gehen. {Er$_{i,j}$/Der$_{i,j}$} sucht sich immer Leute als Trainingspartner aus, die nicht richtig fit sind.

Paul$_i$ wanted to go running with Peter$_j$. He {PPro$_{i,j}$/DPro$_{i,j}$} always picks people as training partners who are not really fit.

In the experiment to be discussed in *The experimental study*, we make use of the anti-logophoricity of DPros as indicators that the sentences containing them have to be interpreted as expressing the speaker's or narrator's perspective (see Zeman, 2019, to appear for discussion of other indicators of the narrator's presence as a perspective taker in narrative texts) rather than the perspective of the discourse referent picked up by the respective DPro. Accordingly, the items which are meant to express the speaker's or narrator's perspective are construed in such a way that they license the use of DPros to pick up the respective topical referent, with the presence of the DPro enforcing such an interpretation. The items which are meant to express the respective topical referent's perspective, in contrast, always contain an instance of FID that can only be interpreted as a thought of that protagonist.

Having discussed perspective taking on the linguistic level, we will discuss perspective taking on the level of co-speech gestures in the following section.

FIGURE 1 | Original scene

Observer Viewpoint and Character Viewpoint Gestures

As already mentioned in the introduction, it is well-known that perspective taking can also be expressed at the level of co-speech gestures, i.e. via gestures that speakers produce while uttering sentences. While there are gestures that are produced without speech and either replace certain parts of speech (pro-speech), precede (pre-speech) or follow it (post-speech), most gestures are produced during speech (co-speech) (see McNeill, 1992 for a descriptive approach of the different types of gestures and Schlenker, 2018 for recent discussion in the formal semantic realm). Empirical studies have shown that a co-speech gesture and the corresponding spoken language segment are not only semantically, but also temporally aligned in systematic ways. Usually the apex (or more generally: the "stroke") of a gesture coincides with or directly precedes an intonational peak, the main accent, of the semantically associated phrase (Pittenger et al., 1960; Kita and Özyürek, 2003; Loehr, 2004). Typically, the content of a speech-accompanying gesture semantically interacts closely with the corresponding phrase and triggers a complex meaning ensemble that is dependent on the utterance context (Kopp et al., 2004). In the following, we will only be concerned with a certain type of gestures among many very different gesture types, namely iconic gestures. Iconic gestures depict some aspect of what they are meant to represent. For example, a "round"-gesture indicating the shape of some round item depicts roundness and thus always bears a certain similarity to roundness or a round object. At the same time, iconic gestures are more or less idiosyncratic and dependent on the person that performs them. While one speaker chooses to illustrate the roundness of a certain object by way of a static two-handed gesture representing roundness, another might use a dynamic gesture drawing a circle in the air with the index finger. While iconic gestures are non-conventionalized and, as their name says, very iconic, these are not properties that are shared by gestures in general. Emblematic gestures, for example, like the "thumbs-up" or the "victory" sign are symbols that have to be performed according to the conventions of a certain cultural community and they do not necessarily bear similarity to what they represent. An emblem can have a certain meaning in one community and a very different one or none in another. In the following, we will only be concerned with iconic gestures.

Crucially, it has been argued that there are two kinds of iconic gestures which are often used by speakers when they describe scenes or events to their interlocuters and that clearly reveal a perspective: character viewpoint gestures (CVG), on the one hand, and observer viewpoint gestures (OVG) gestures, on the other (McNeill, 1992; Parrill, 2010, Parrill, 2012; Stec, 2012, Stec, 2016). When producing a CVG, the speaker impersonates an individual participating in the event described by the sentence and enacts the event from that person's point of view. When producing an OVG, in contrast, the speaker depicts the event described by the sentence as if it was observed from a distance. CVGs typically involve the speaker's entire body and face, while OVGs are usually performed exclusively with the hands (McNeill 1992).

As an illustration, consider the two different gesture types in **Figures 1**–**3** from Parrill (2010: 651). She discusses two gestures performed by participants in an experiment in which they had to describe to their interlocuters a scene from a cartoon they had just seen: an event of a skunk hopping across the room. In one case, the speaker performed an OVG in which the hand represents the skunk and the trajectory of the hand represents the trajectory of the skunk. Crucially, the event was depicted as if observed from a distance and in a rather schematic way. In the other case, in contrast, the speaker performed a CVG in which he enacted the hopping skunk with his entire body.

FIGURE 2 | Observer Viewpoint Gestures

The Relationship Between Linguistic and Gestural Viewpoint

We follow De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; see also Kendon, 2004) in assuming that combinations of speech and gesture convey a complex multimodal message that is planned by a central cognitive process and then dispatched into disparate channels. The information conveyed via the two channels may be either redundant (thus avoiding misunderstanding) or complementary. Although Goldin-Meadow. (1999) has shown that in cases of learning novel concepts that involve transition between different cognitive states, mismatches between gesture and speech can be quite productive and useful, the default should be for the information conveyed via one channel to be coherent with information conveyed via the other one. If Parrill (2010) is right in assuming that gestural and linguistic viewpoint have a common conceptual source, perspective is an integral part of the multimodal message conveyed by the combination of speech and gesture. Consequently, the default should be for such a message to be coherent with regard to perspective as well, i.e. it should express a single perspective by default. Concerning perspective, this means that at least in the absence of intervening factors there should be a preference for perspective alignment. This should concern production as well as comprehension, i.e. the speaker should in the default case plan to convey a message that is coherent with regard to perspective and the listener should in the default case expect the speaker to convey such a message.

There is, however, the complicating case of multiple perspectives. It is possible that a speaker plans to convey a thought from more than one perspective. Although there is some discussion about the possibilities of linguistic realizations of such multiple perspectives within the evidentiality literature (see Evans, 2005 and Bergqvist, 2015 for some discussion), a systematic investigation of the linguistic tools to simultaneously convey multiple perspectives is still outstanding. As for the gestural realization of perspective, McNeill (1992), Cassell et al. (1999) and Parrill (2009) discuss exactly such examples of dual viewpoints. Cassell et al. (1999) present an example where someone hands something to himself while uttering she got something. Here, the arm and hand embodies the giver and the rest of the body the receiver. McNeill (1992) and Parrill (2009) discuss similar cases as well as cases where someone performs an OVG and a CVG at the same time. Parrill (2009) reports an example where the narrator talks about a character and impersonates the reported character by performing a body lean to mimic the action of this character, hence a CVG, and at the same time indicates the trajectory of a certain path taken by the character, clearly an OVG.

So far, the relation between linguistic and gestural perspective has not been systematically investigated. If, however, there are techniques to represent dual perspectives within speech alone and within gesture alone, as we have pointed out above, it is not implausible to assume that a speaker can also choose to convey two perspectives at the same time and realize one via gesture and one via speech. We believe, however, it is equally fair to assume that such dual viewpoint realizations need specific licensing conditions and are the exception rather than the general case. Future research will have to shed light on this.

Existing research on the relationship of viewpoint gestures and speech has for the most part focused on general factors influencing the frequency with which co-speech gestures are performed or which types of gestures are preferred, depending on the accompanying speech. McNeill (1992), for example,

FIGURE 3 | Character Viewpoint Gesture.

observed that linguistic complexity influences which type of co-speech gesture is performed, with utterances containing transitive verbs having a tendency to be accompanied by CVGs and sentences with intransitive verbs having a tendency to be accompanied by OVGs. Additionally McNeill (1992), notes a tendency for causally central events to be accompanied by CVGs. The first observation was confirmed by Parrill (2010), while the second was disconfirmed: While causally central events were more often accompanied by gestures than peripheral events, there was no contrast between CVGs and OVGs. At the same time, Parrill (2010), Parrill (2012) found other factors that had an influence on which type of gesture was chosen by speakers. First, she found that speakers performed CVGs more often than OVGs when the information conveyed by the utterance was new to the hearer, while for shared information it was the other way around. Second, she found that the internal structure of the reported event was a crucial factor, too, with events involving the display of affects or a prominent use of the character's hands and torso triggering more CVGs than OVGs, while for events involving trajectories it was the other way around.

Having discussed perspective taking on the linguistic level, on the level of co-speech gestures, as well as their relationship, we will discuss an experimental study in which we systematically investigated the interaction of the two kinds of perspective taking in light of our assumptions concerning perspectival coherence in *The experimental study* (see Ebert and Hinterwimmer to appear for a study of self-pointing CVGs in reported speech vs. direct speech vs. FID and a proposal to account for such and other demonstrations in quotation on basis of Ebert et al. (2020) account for co-speech gestures).

THE EXPERIMENTAL STUDY

As stated above, we assume that

(a) Gesture and speech together convey a multimodal message that is planned by a central cognitive process and then dispatched into disparate channels (De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; see also; Kendon, 2004).
(b) Gestural and linguistic viewpoint have the same conceptual source (Parrill, 2010).
(c) Perspective is an integral part of the multimodal message to be conveyed and,
(d) the default is for this message to be coherent.

We thus predict a strong preference for gestural and linguistic perspective to be aligned in a single utterance, at least in the absence of intervening factors.

In order to test this prediction, we conducted a forced-choice experiment in German in which we tested whether utterances in which the linguistically expressed perspective aligns with the perspective expressed on the gestural level are preferred to test items in which this is not the case. We compared variants of test items clearly expressing the perspective of an individual participating in the events described with variants of test items which clearly express the speaker's or narrator's perspective. Variants of the former kind were always construed in such a way that the opening sentence describes the feelings, intentions or thoughts of a protagonist in a particular situation which the second sentence specifies in more detail, while the final sentence renders a thought that the respective protagonist has in that situation in the form of FID. Variants of the latter kind, in contrast, always contained an opening sentence in the form of a general statement in present tense expressing an evaluation of or opinion about the individual participating in the event described by the following sentence. Additionally, that individual was referred to by a DPro. The two variants of each test item basically described the same situation. Each variant of each test item was videotaped in two different versions: In one version, the speaker performed a CVG while uttering the respective final sentence. In the second version, the speaker performed an OVG. Two examples are provided in (9) and (10). The respective CVGs and OVGs, which are described beneath the items, were performed while uttering the portion of the respective sentence marked in boldface. All stimuli, anonymized data, and codes can be accessed via the following link: https://osf.io/4bqpx/.

(9) a. Leon ist ein begeisterter Sportler. Als der sich neulich beim Fußballspielen den Ball erkämpfte, **kickte er ihn sofort in Richtung Tor (narrator perspective = NP)**.
Leon is an enthusiastic athlete. When he (DPro) recently won the ball while playing soccer, he immediately kicked it in the direction of the goal.
b. Leon spielte am Wochenende Fußball. Nach einigem Gerangel hatte er sich den Ball erkämpft. **Toll, jetzt konnte er ihn direkt in Richtung Tor schießen (character perspective = CP)**
Leon played soccer on the weekend. After some scramble, he had finally won the ball. Great, now he could directly kick it in the direction of the goal!
CVG: The speaker performs a kicking movement with her right leg and foot, displaying an enthusiastic facial expression.
OVG: The speaker presses her index finger on her thumb, then releasing it quickly, thus imitating a kicking movement with the index finger.

(10) a. Denise ist ein richtiger Tollpatsch. Als die neulich nach Feierabend das Büro verließ, hat sie nicht richtig aufgepasst **und ist voll gegen die Tür geknallt! (NP)**
Denise is a real klutz. When she (DPro) recently left the office at the end of the work day, she did not really pay attention and slammed into the door at full tilt.
b. Denise hatte es eilig. Beim Verlassen des Büros passte sie nicht richtig auf. **Autsch, jetzt war sie voll gegen die Tür geknallt (CP).**
Denise was in a hurry. When leaving the office, she did not really pay attention. Ouch, now she had slammed into the door at full tilt!

CVG: The speaker throws back her head and imitates the astonished facial expression of someone banging her head against a door, eyes wide open.
OVG: The speaker moves the index finger of her right hand quickly into the direction of her vertically upheld left hand and lets it collide with her hand and bounce back.

The variants of the test items were evenly distributed across two lists, so that each participant saw only either the NP or the CP variant of each test item in both conditions, i.e. in the version where the speaker performs a CVG while uttering it and in the version where she performs an OVG while uttering it. The participants then had to choose the video with the gesture that they thought better fits the spoken utterance. The method we employed is thus similar to the one employed in grammaticality and/or pragmatic felicity judgement tasks, the underlying reasoning being that speakers have intuitions regarding the interaction of linguistic and gestural perspective taking that reflect underlying unconscious principles in the same way as they have intuitions reflecting unconscious syntactic, semantic or pragmatic principles.

Now, if there is a strong preference for linguistic and gestural perspective to be aligned, the CVG version should be chosen more often for the CP variants of the test items, while the OVG should be chosen more often for the NP version of the test items. In (9b), for example, the sentence in boldface renders a thought of Leon that expresses Leon's perspective on the event of him kicking the ball in the direction of the goal. When performing the CVG described above while uttering that sentence, the speaker enacts the event of Paul kicking the ball in the direction of the goal from his perspective. Consequently, the linguistically and the gesturally expressed perspectives align when (9b) is combined with the CVG. When performing the OVG described above while uttering the sentence in boldface in (9b), in contrast, she depicts the event of Leon kicking the ball in the direction of the goal from an outside perspective. There is thus a mismatch between the linguistically and the gesturally expressed perspective.

In (9a), the opening sentence in combination with the use of the DPro to refer to Leon in the temporal adjunct clause ensures that the sentence in boldface is attributed to the narrator or speaker, i.e. it reports the event of Leon kicking the ball in the direction of the goal from the narrator's or speaker's perspective. When the speaker performs the OVG described above while uttering the sentence in boldface in (9a), the linguistically expressed perspective aligns with the gesturally expressed one, since the outside perspective conveyed by the OVG can easily be construed as the narrator's or speaker's perspective. When she performs the CVG while uttering the sentence in (9a), in contrast, there is a mismatch between the linguistically expressed narrator's or speaker's perspective and the gesturally expressed perspective, which is Leon's perspective.

Consequently, if there is a requirement or a strong preference for the linguistically and the gesturally expressed perspective to be aligned (9a) should be chosen more often in combination with the OVG, and (9b) in combination with the CVG. At the same time, as we have seen in *Observer viewpoint and character viewpoint gestures* above, there is a preference for complex events (i.e. events involving at least two participants) as well as for events contained in sentences introducing new information to be accompanied by CVGs. The sentences in boldface in (9) and (10) clearly convey information that is new to the participants and they contain transitive verbs. Consequently, if there is no requirement or strong preference for the linguistically expressed perspective to align with the gesturally expressed perspective, CVGs should be preferred across both conditions according to the above-mentioned findings of Parrill (2010, 2012).

In order to make sure that our stimuli were actually interpreted as intended, i.e. as either expressing the character's or the narrator's perspective, we conducted an informal forced-choice study[1]. Participants saw muted versions of both the CVG- and the OVG-variant of each test item together with a short, neutral description of the situation reported by both variants of the respective item. For (9a–b), for example, the following description was provided: *The following video is about a soccer player who is kicking the ball into the goal.* Participants (n = 18) then had to decide for each item which of the two videos corresponds to the character's and which one to an observer's perspective, i.e., in effect, whether the gesture they saw was a CVG or an OVG[2]. Additionally, they had to indicate on a scale from 1 to 5 how sure they were of their judgement, with 1 expressing minimal and 5 expressing maximal confidence. In terms of the design, the experimental task was almost equivalent to showing both video versions (CharVideo and ObsVideo) of an item and asking participants to choose which one is more likely to be a CVG and which one is more likely to be an OVG (i.e. it was almost equivalent to a classical forced-choice design)[3].

As **Figure 4**, Panel **(A)** shows, in the overwhelming majority of cases, the gestures were interpreted as intended, i.e. videos with a CVG were interpreted as conveying the character's perspective and videos with an OVG as conveying an observer's perspective. Additionally, as Panel **(B)** shows, in the vast majority of cases, participants were confident in their choices, i.e. there were very few low confidence responses (i.e. only 14% of all responses were given with confidence lower than 3).

We now turn to a detailed description and discussion of the experimental study itself.

[1]We are grateful to two anonymous reviewers as well as the editor for urging us to conduct this study.

[2]Participants were recruited through Prolific (https://prolific.ac/) for monetary compensation (3,75 £; 7,50 £/h), just as in the main experiment. Only persons who had not participated in the main experiment could participate. Two participants had to be excluded because they self-identified as non-native speakers of German. The study also included simple questions which were included in order to check whether participants were paying attention. No one had to be excluded on the basis of answering these questions incorrectly.

[3]We chose that design in order to keep the study similar to the design of the main experiment.

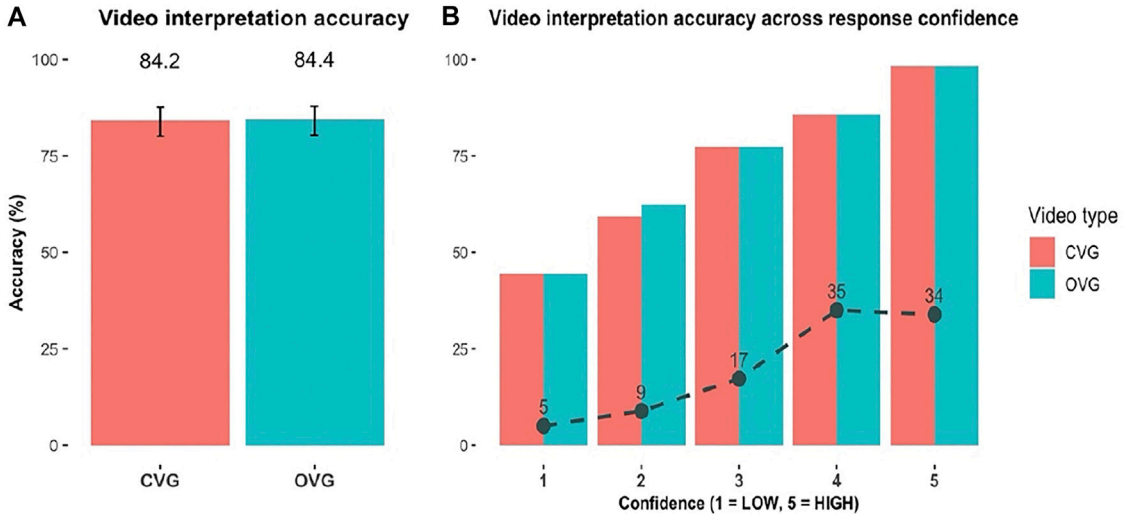

FIGURE 4 | The bar plots show mean percentages of accurately recognizing CVG and OVG videos. Panel **(A)** shows the mean accuracy for the two types of videos (the confidence intervals are calculated using prop.test() function in R). Panel **(B)** shows accuracy across five levels of response confidence with "1" denoting least and "5" most confident responses. The overplotted dots joined with a dashed line in Panel **(B)** show percentages of responses in each level of the confidence scale (the numbers above the dots are the percentages).

METHOD

Participants

Eighty-five native speakers of German (45 males, 40 females, age 18–65) were recruited through Prolific (https://prolific.ac/) for monetary compensation (3,75 £; 7,50 £/h).

MATERIALS

We constructed 20 experimental items similar to those in (9) and (10), interspersed with 24 fillers. For each item there was an NP variant (similar to (9a) and (10a)) and a CP variant (similar to (9b) and (10b)). Both variants always described the same situation. Each variant of each test item was videotaped in two different versions: In one version, the speaker performed a CVG while uttering the respective final sentence. In the second version, the speaker performed an OVG. Consequently, each item came in four different conditions: NP-CVG, NP-OVG, CP-CVG and CP-OVG.

The fillers consisted of two sentences and involved pointing to a location in the gesture space. In the first sentence, two discourse referents were introduced by referential expressions accompanied by pointing gestures that anchored the referents in the gesture space. In 12 filler items both referents had the same gender (e.g. *Gestern auf der Party hat Peter Linus beleidigt.* Engl.: *Yesterday at the party, Peter insulted Linus.*, plus pointing to a point left in the central gesture space in front of the speaker's body when uttering Peter and to a point right when uttering Linus.), and in 12 filler items the gender was different (e.g. *Gestern hat Martin Claudia zum Abendessen eingeladen.* Engl.: *Yesterday Martin invited Claudia for dinner.*, plus pointing to a point left in the central gesture space in front of the speaker's body when uttering *Martin* and to a point right when uttering *Claudia.*). The second sentence always contained a pronoun and the speaker pointed to the location associated with the object referent while uttering the pronoun. For each filler item there were two different versions. For the 12 filler items where both referents had the same gender there was a version with a DPro and a version with a PPro (e.g. *Der hat dann sofort angefangen zu weinen.* Engl.: *He (DPro) then started crying immediately*, and *Er hat dann sofort angefangen zu weinen*. Engl.: *He then started crying immediately*). For the 12 filler items where the gender was different for the two referents there was a version with a male DPro and a version with a female DPro (e.g. *Der hat sich sehr darüber gefreut.* Engl.: *He (DPro) was very happy about it* and *Die hat sich sehr darüber gefreut.* Engl.: *She (DPro) was very happy about it*).

Procedure

The experiment, which involved a forced-choice task, was conducted online. The NP- and the CP-variants of the test items were evenly distributed across two lists and presented in pseudo-randomized order, interspersed with the fillers. Consequently, participants saw either both the NP-OVG and the NP-CVG version of the respective test item or the CP-CVG and the CP-OVG version. Concerning the filler items, they always saw both versions, i.e. the versions with the DPro and the PPro and the versions with the DPro matching the gender of the referent associated with the location pointed at and the version not matching it.

The task for the participants was to choose the version of the test items or the fillers in which the combination of language and gesture is more natural according to their intuitions. The question that appeared below the video was: *Welche Geste passt besser zur*

sprachlichen Äußerung in den Videos? Engl.: *Which gesture better fits the spoken utterance in the videos?*). Participants were told beforehand to pay good attention to sound and picture and that they had the option to replay the videos. They were also told that in some cases the utterances were identical in the two versions and the gestures accompanying them were different (as in the test items), while in other cases it was the other way around (as in the filler items). Before the experiment started, the participants were shown two trial items to familiarize them with the form of the experiment. They also had to do four simple matching tasks, the purpose of which was to check whether they were paying attention: They were shown two pictures of animals, fruits etc. and had to decide which of the two pictures matched a word such as *cat*, *apple* etc. Participants who would choose the wrong picture in one of the four tasks were to be excluded from the final analysis (however, there was no such case).

Data Analysis

We excluded seven participants from the analysis because they either did not complete the task or completed it in less than 10 min (the approximate duration was 30 min). All data processing and analyses were carried out in R (Core Team, 2020). Since the responses were binomial (CVG or OVG), we analyzed the proportions of CVG responses using mixed-effects logistic regression through the R package brms (Bürkner, 2017). We used condition as the predictor variable with CP as the reference level. To avoid the extreme probability values (0 and 1), we used weakly informative priors, N(0, 2.5), instead of the default priors of brms for logistic regression (Student-*t* with df = 3, mu = 0 and sigma = 2.5). The model was run with four sampling chains each of which ran for 5,000 iterations with a warm-up period of 2000 iterations. We also fit another model with the same form but without any predictors, an intercept only model, to statistically test if the CVG option was chosen more often than the OVG option.

For each effect we report its mean and 95% CrI under the posterior distribution. We use CrI to make inferences about the presence of an effect. If the 95% CrI for an effect does not include zero we consider that there is compelling evidence for that effect. We also report the posterior probability of an effect being greater than zero or less than zero depending on the sign of the estimated parameter mean. The posterior probability is calculated using the posterior sample for a parameter generated by the statistical model and it is the proportion of the sample less than or greater than zero.

Results

The response proportions are plotted in **Figure 5**, and the results of the data analysis are listed in **Table 1** for the model with condition as the predictor and **Table 2** for the intercept only model. We found that overall the CVG option was chosen clearly more often than the OVG option (**Table 2**). Moreover, the proportions of CVG responses were influenced by the condition (CP vs. NP) such that in the NP condition the CVG option was chosen less often than in the CP condition; although this effect was small and the support for it was weaker (**Table 1**).

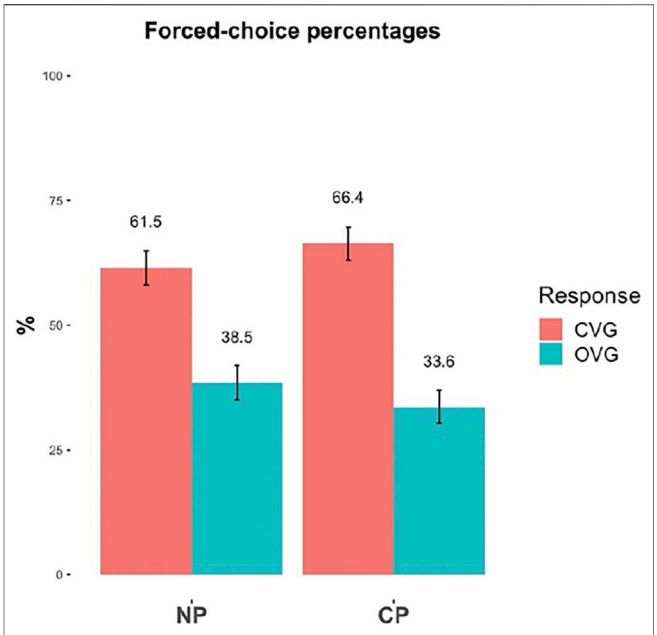

FIGURE 5 | Each bar shows the percentage of choosing an option (CVG or OVG) across the two conditions, NP and CP (the confidence intervals are calculated using prop.test() function in R).

TABLE 1 | Results from statistical analysis—estimates of the model, corresponding 95% credible intervals (95% CrI) and the posterior probabilities (Post. Prob.). Effects for which the CrI excludes zero are shown in bold. 'Intercept' denotes the effect of selecting CVG more often than OVG in the CP condition, and Narrator denotes the effect of selecting CVG in the Narrator condition compared to the CP.

Effect	Estimate	95% CrI	Post. prob
Intercept	1.10	[0.35, 1.89]	0.997
Narrator	−0.32	[−0.75, 0.12]	0.926

TABLE 2 | Results from statistical analysis—estimates of the model, corresponding 95% credible intervals (95% CrI) and the posterior probabilities (Post. Prob.). Effects for which the CrI excludes zero are shown in bold. In this intercept-only model 'Intercept' denotes the effect of selecting CVG more often than OVG.

Effect	Estimate	95% CrI	Post. prob
Intercept	0.90	[0.20, 1.63]	0.993

DISCUSSION

At first sight, the experimental results seem to be incompatible with the assumption of a strong preference for the linguistically expressed perspective to align with the perspective expressed on the level of co-speech gestures. If such a strong preference existed, participants should have chosen the combination of CP and CVG more often than the combination of CP and OVG, on the one hand, and the combination of NP and OVG more often than the

combination of NP and CVG, on the other. Rather, they showed a clear preference for CVGs across both the CP and the NP condition. This is in line with the findings of McNeill (1992) and Parrill (2010), Parrill (2012) that sentences with transitive verbs and sentences introducing new information have a tendency to be accompanied by CVG since the relevant portions of our test items all introduced information that was new to the participants and the vast majority of them contained transitive sentences.

Although the preference for CVGs was slightly stronger in the CP condition than in the NP condition, this preference did not turn out to be significant, so no conclusions can be derived from it at this point. Nevertheless, there is still the possibility that linguistic and gestural perspective are preferably aligned, but that this is a constraint that ranks below the requirement that sentences introducing new information and describing rather complex events should be accompanied by CVGs. Concerning the question of why the two latter preferences exist, we would like to tentatively suggest that this is due to CVGs always being more informative than OVGs, which only depict events in a rather generic and schematic way. CVGs, in contrast, by making use of the speaker's entire body in combination with her facial expression, convey much more fine-grained and detailed information which is particularly useful when the sentence they accompany introduces new information and when the event described by that sentence is rather complex.

Recall that we derived our hypothesis of a strong preference for gestural and linguistic perspective to be aligned in a single utterance from the following assumptions:

(a) Gesture and speech together convey a multimodal message that is planned by a central cognitive process and then dispatched into disparate channels (De Ruiter, 1998, De Ruiter, 2000, De Ruiter, 2007; see also Kendon, 2004).
(b) Gestural and linguistic viewpoint have the same conceptual source (Parrill, 2010).
(c) Perspective is an integral part of the multimodal message to be conveyed.
(d) The default is for this message to be coherent.

Since we still consider these assumptions to be very plausible, the most straightforward way to reconcile them with the findings of our study would be to assume the preference for informative gestures to be strong enough to overwrite the default. At the same time, the differences between the OVGs and the CVGs in our study can be interpreted as revealing a potential limitation, which might have affected the results[4]. After all, the co-speech gestures did not only differ with respect to perspective and informativity, but also with respect to size, since the CVGs involved the speaker's entire body and facial expression, while OVGs only involved the hands. This difference in size quite plausibly made the CVGs more salient than the OVGs, which might have played a role in the general preference for CVGs. Additionally, CVGs might have been judged as more natural in virtue of the speaker being more fully engaged. One might consider to conducting a follow-up study in which the gestures are more comparable in size and speaker's engagement and therefore in saliency and naturalness. To give a concrete example, the OVG in (9a–b), repeated here as (11a–b), could be replaced by a full-body gesture where the speaker steps back and follows the path of an imaginary ball with her index finger and gaze. This would, however, mean departing from the standard view that OVGs are performed with only the hands. It might hence be more feasible to replace the CVG by a CVG where the gesturer does only a small kick of the foot but does not incorporate her upper body or facial expression.

(11) a. Leon ist ein begeisterter Sportler. Als der sich neulich beim Fußballspielen den Ball erkämpfte, **kickte er ihn sofort in Richtung Tor. (narrator perspective = NP).**
Leon is an enthusiastic athlete. When he (DPro) recently won the ball while playing soccer, he immediately kicked it in the direction of the goal.
b. Leon spielte am Wochenende Fußball. Nach einigem Gerangel hatte er sich den Ball erkämpft. **Toll, jetzt konnte er ihn direkt in Richtung Tor schießen. (character perspective = CP).**
Leon played soccer on the weekend. After some scramble, he had finally won the ball. Great, now he could directly kick it in the direction of the goal!
CVG: The speaker performs a kicking movement with her right leg and foot, displaying an enthusiastic facial expression.
OVG: The speaker presses her index finger on her thumb, then releasing it quickly, thus imitating a kicking movement with the index finger.

We are planning to conduct a follow-up study with gestures that are more comparable in saliency and naturalness in order to test whether the difference in size between the two kinds of gestures in our original study had an influence on the results. Additionally, in virtue of the preference for CVGs being potentially linked to the speaker's introducing new information, we are planning to conduct a follow-up study which does not only contain stimuli with new events, as the study reported in this paper does, but which contains both stimuli with new and stimuli with given events. Our prediction is that the preference for CVGs should at least be weaker in the stimuli with given events and potentially be overwritten by the preference for perspective alignment.

Let us finally add a grain of salt: Since the division of labour between gesture and speech in general and viewpoint issues in particular have not been systematically investigated and are not settled yet, we could not control for potential intervening factors that might elicit multiple viewpoint representations and a potential split of viewpoints on gesture and speech. As it stands, our results are equally compatible with the possibility that the gestural channel preferably transports the character

[4]We thank an anonymous reviewer for pointing this out to us.

perspective, independent of the perspective that the speech channel transports.

CONCLUSION

The topic of this paper was an investigation of the interaction of perspective taking expressed on the linguistic level with perspective taking expressed on the level of co-speech gestures. We investigated via a forced-choice experimental study whether there is a preference for the linguistic perspective to be aligned with the gestural perspective. The experimental results provided no evidence that there is such a preference for perspective alignment. There was a general preference for CVGs, which was slightly, but not significantly, stronger when the respective CVG accompanied a sentence expressing a character's perspective, however, than when it accompanied a sentence expressing the narrator's or speaker's perspective. After all, the results of our study did not support our initial hypothesis that there is a preference for perspective alignment. It might, however, still be the case that there is such a preference and it can be overwritten by the preference for more informative gestures or by a preference to transport the character's perspective in the gestural channel. In future research we are planning to conduct studies which are aimed at testing for this possibility. First, we are planning a follow-up study with CVGs and OVGS that are more comparable in saliency and naturalness than those in our original study in order to test whether the differences in saliency and naturalness were a confounding factor. Secondly, we are planning to conduct a follow-up study which contains both stimuli with new and stimuli with given events in order to test whether the preference for CVGs in our original study was at least partly due to CVGs being preferred in sentences conveying new information. Finally, we are planning a study in which the following two cases will be compared:

a. Utterances where the speaker describes an event she participated in from her first person perspective while performing a CVG that depicts the actions of another participant.
b. Utterances where the speaker describes an event she participated in from her first person perspective while performing a CVG that depicts her own actions.

If our assumptions are correct, utterances instantiating the constellation in a. should be clearly dispreferred compared to utterances instantiating the constellation in b. since the former in contrast to the latter involve conflicting perspectives. At the same time, they do not differ in informativity, since the gestures in both cases are CVGs. If there are no clear differences between the two cases, this would be a strong indication that there is not even a weak default preference for linguistic and gestural perspective to be aligned. Since only CVGs and no OVGs are involved, a gestural preference for CVGs should not influence the results, either.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/4bqpx/.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SH and CE jointly developed the experimental design and the analysis and were jointly responsible for the execution of the experiment, UP was responsible for the statistical analysis. The introduction and the parts on Free Indirect Discourse and DPros were written by SH alone, the section on gestures by CE, and the sections discussing the experiment, the discussion and the conclusion were jointly written by SH, CE and UP.

FUNDING

The research reported in this paper was funded by the DFG 1459 within the SFB 1252 Prominence in Language (Project-ID 281511265) (SH) and as part of the project PSIMS of the Priority Program 1727 XPRAG. de (EB 523/1-1) (CE).

ACKNOWLEDGMENTS

We thankfully acknowledge the support of the German Research Foundation (DFG). We would like to thank Theresa Stender, Sebastian Walter, Janne Schmandt, Felix Jüstel and Magdalena Schmitz for help with item generation and execution of the forced-choice study reported in *The Experimental Study*. We also thank Gregor Brinkmeier from studiumdigitale of the University of Frankfurt for supplying us with a studio and all technical equipment for the recordings as well as help with the postprocessing of the recordings.

REFERENCES

Abrusán, M. (2020). The Spectrum of Perspective Shift: Protagonist Projection vs. Free Indirect Discourse. *Linguistics And Philosophy*. *Adv. access published July* 22. doi:10.1007/s10988-020-09300-z

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston: Routledge & Kegan Paul.

Bergqvist, H. (2015). Epistemic Marking and Multiple Perspective: An Introduction. *Lang. Typology Universals* 68 (2), 123–141. doi:10.1515/stuf-2015-0007

Bosch, P., Katz, G., and Umbach, C. (2007). "The Non-subject Bias of German Demonstrative Pronouns,". Editors M. Schwarz-Friesel, M. Consten, and M. Knees, 145–164. doi:10.1075/slcs.86.13bos Anaphors in textAmsterdam & Philadelphia: Benjamins.

Bosch, P., and Umbach, C. (2006). Reference Determination for Demonstrative Pronouns. Proceedings of the conference on intersentential pronominal reference in child and adult language. Editors D. Bittner and N. Gagarina (Berlin: Zentrum für Allgemeine Sprachwissenschaft, Sprachtypologie und Universalforschung), 48, 39–51.

Brinton, L. (1980). 'Represented Perception': A Study in Narrative Style. *Poetics* 9, 363–381. doi:10.1016/0304-422x(80)90028-5

Bürkner, P.-C (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw. Articles* 80 (1), 1–28. doi:10.18637/jss.v080.i01

Cassell, J., McNeill, D., and McCullough, K.-E. (1999). Speech-gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Nonlinguistic Information. *P&C* 7 (1), 1–34. doi:10.1075/pc.7.1.03cas

Charnavel, I., and Mateu, V. (2015). The Clitic Binding Restriction Revisited: Evidence for Antilogophoricity. *Linguistic Rev.* 32 (4), 671–701. doi:10.1515/tlr-2015-0007

Clements, G. N. (1975). The Logophoric Pronoun in Ewe: Its Role in Discourse. *The J. West African Lang.* 10, 141–177. doi:10.17487/rfc0683

De Ruiter, J. P. (1998). Gesture and Speech Production, PhD Thesis. Nijmegen, Netherlands: University of Nijmegen.

De Ruiter, J. P. (2007). Postcards from the Mind. *Gest* 7, 21–38. doi:10.1075/gest.7.1.03rui

De Ruiter, J. P. (2000). "The Production of Gesture and Speech," in *Language and Gesture*. Editor D. McNeill (Cambridge, UK: Cambridge University Press), 284–311. doi:10.1017/cbo9780511620850.018

Dirscherl, F., and Pafel, J. (2015). Die vier Arten der Rede- und Gedankendarstellung. Zwischen Zitieren und Referieren. *Linguistische Berichte* 241, 3–47.

Doron, E. (1991). "Point of View as a Factor of Content," in *Proceedings of Semantics and Linguistic Theory*. Editors S. K. Moore and A. Z. Wyner (Ithaca, NY: Cornell University), 1, 51–64. doi:10.3765/salt.v1i0.2997

Ebert, C., and Hinterwimmer, S. (2020). "Free Indirect Discourse Meets Character Viewpoint Gestures: A Reconstruction of Davidson's Demonstration Account with Gesture Semantics," in To Appear appear Linguistic Evidence 2020 Proceedings.

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

Eckardt, R. (2014). *The Semantics of Free Indirect Discourse. How Texts Allow to Mind-Read and Eavesdrop*. Leiden: Brill.

Evans, N. (2005). View with a View: Towards a Typology of Multiple Perspective Constructions. *Bls* 31, 93–120. doi:10.3765/bls.v31i1.3429

Goldin-Meadow, S. (1999). The Role of Gesture in Communication and Thinking. *Trends Cogn. Sci.* 3, 419–429. doi:10.1016/s1364-6613(99)01397-2

Hinterwimmer, S. (2015). "A Unified Account of the Properties of German Demonstrative Pronouns," in The proceedings of the workshop on pronominal semantics at NELS. Editors P. Grosz, P. Patel-Grosz, and I. Yanovich (Amherst, MA: GLSA Publications, University of Massachusetts), 61–107.

Hinterwimmer, S. (2020). Zum Zusammenspiel von Erzähler- und Protagonistenperspektive in den Brenner-Romanen von Wolf Haas," in *Sprachlichen Strukturen der Narration, special issue of Zeitschrift für Germanistische Linguistik*. Editor S. Zeman ((ZGL)). doi:10.1515/zgl-2020-2013

Hinterwimmer, S., and Bosch, P. (2016). "Demonstrative Pronouns and Perspective," in *The Impact of Pronominal Form on Interpretation*. Editors P. Patel and P. Patel-Grosz (Berlin/New York: De Gruyter (Studies in Generative Grammar)).

Hinterwimmer, S., and Bosch, P. (2017). "Demonstrative Pronouns and Propositional Attitudes," in *Pronouns in Embedded Contexts (Studies in Linguistics and Philosophy)*. Editors P. Patel-Grosz, P. G. Grosz, and S. Zobel (Washington, DC: Springer), 282–301. doi:10.1007/978-3-319-56706-8_4

Hinterwimmer, S., and Brocher, A. (2018). An Experimental Investigation of the Binding Options of Demonstrative Pronouns in German. *Glossa: A J. Gen. Linguistics* 3 (1), 77. doi:10.5334/gjgl.150

Hinterwimmer, S., Brocher, A., and Patil, U. (2020). Demonstrative Pronouns as Anti-logophoric Pronouns: An Experimental Investigation. *Dialogue Discourse* 11 (2), 110–127. doi:10.5087/dad.2020.204

Hinterwimmer, S., and Meuser, S. (2019). , "Erlebte Rede und Protagonistenprominenz,". *Rede- und Gedankenwiedergabe in narrativen Strukturen – Ambiguitäten und Varianz, Linguistische Berichte, Sonderheft*. Editors S. Engelberg, C. Fortmann, and I. Rapp (Hamburg: Helmut Buske Verlag), 27, 177–200.

Hinterwimmer, S. (2019). Prominent Protagonists. *J. Pragmatics* 154, 79–91. doi:10.1016/j.pragma.2017.12.003

Hinterwimmer, S. (2017). Two Kinds of Perspective Taking in Narrative Texts, *Salt* In D. Burgdorf, J. Collard, S. Maspong, and B. Stefánsdóttir (eds.), Proceedings of Semantics and Linguistic Theory(SALT) 27, 282–301. doi:10.3765/salt.v27i0.4153

Kaiser, E. (2013). Looking beyond Personal Pronouns and beyond English: Typological and Computational Complexity in Reference Resolution. *Theor. Linguistics* 39, 109–122. doi:10.1515/tl-2013-0007

Kaiser, E. (2011a). "On the Relation between Coherence Relations and Anaphoric Demonstratives in German," in *Proceedings of Sinn und Bedeutung*. Editors I. Reich, E. Horch, and D. Pauly (saarbrücken, Germany: Saarland University Press), 15, 337–351.

Kaiser, E. (2010). Effects of Contrast on Referential Form: Investigating the Distinction between strong and Weak Pronouns. *Discourse Process.* 47, 480–509. doi:10.1080/01638530903347643

Kaiser, E. (2011b). Salience and Contrast Effects in Reference Resolution: The Interpretation of Dutch Pronouns and Demonstratives. *Lang. Cogn. Process.* 26 (10), 1587–1624. doi:10.1080/01690965.2010.522915

Kaiser, E., and Trueswell, J. C. (2008). Interpreting Pronouns and Demonstratives in Finnish: Evidence for a Form-specific Approach to Reference Resolution. *Lang. Cogn. Process.* 23 (5), 709–748. doi:10.1080/01690960701771220

Kaplan, D. (1989). Demonstratives. in *Themes from Kaplan*. Editors J. Almog, J. Perry, and H. Wettstein (Oxford: Oxford University Press), 565–614.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University. doi:10.1017/cbo9780511807572

Kita, S., and Özyürek, A. (2003). What Does Cross-Linguistic Variation in Semantic Coordination of Speech and Gesture Reveal?: Evidence for an Interface Representation of Spatial Thinking and Speaking. *J. Mem. Lang.* 48 (1), 16–32. doi:10.1016/s0749-596x(02)00505-3

Kopp, S., Tepper, P., and Cassell, J. (2004). Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In Proceedings of the 6th International Conference on Multimodal Interfaces, 97–104. doi:10.1145/1027933.1027952

Loehr, D. (2004). Gesture and Intonation. PhD thesis. Washington, DC: Georgetown University.

Maier, E. (2015). Quotation and Unquotation in Free Indirect Discourse. *Mind Lang.* 30 (3), 345–373. doi:10.1111/mila.12083

Maier, E. (2017). "The Pragmatics of Attraction," in *The Semantics and Pragmatics of Quotation*. Editors P. Saka and M. Johnson (Dordrecht: Springer), 259–280. doi:10.1007/978-3-319-68747-6_9

Mayol, L., and Clark, R. (2010). Pronouns in Catalan: Games of Partial Information and the Use of Linguistic Resources. *J. Pragmatics* 42, 781–799. doi:10.1016/j.pragma.2009.07.004

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

Nishigauchi, T. (2014). Reflexive Binding: Awareness and Empathy from a Syntactic point of View. *J. East. Asian Linguist* 23, 157–206. doi:10.1007/s10831-013-9110-6

Parrill, F. (2012). Interactions between Discourse Status and Viewpoint in Co-speech Gesture. in *Viewpoint in Language: A Multimodal Perspective*. Editors B. Dancygier and E. Sweetser (Cambridge: CUP), 97–112.

Parrill, F. (2009). Dual Viewpoint Gestures. *Gest* 9 (3), 271–289. doi:10.1075/gest.9.3.01par

Parrill, F. (2010). Viewpoint in Speech-Gesture Integration: Linguistic Structure, Discourse Structure, and Event Structure. *Lang. Cogn. Process.* 25 (5), 650–668. doi:10.1080/01690960903424248

Patil, U., Bosch, P., and Hinterwimmer, S. (2019). Constraints on German Diese Demonstratives: Language Formality and Subject-Avoidance. *Glossa: A J. Gen. Linguistics* 5 (1), 1–22. doi:10.5334/gjgl.962

Pearson, H. (2015). The Interpretation of the Logophoric Pronoun in Ewe. *Nat. Lang. Semantics* 23, 77–118. doi:10.1007/s11050-015-9112-1

Pittenger, R. E., Hockett, C. F., and Danehey, J. J. (1960). *The First Five Minutes: A Sample of Microscopic Interview Analysis*. New York: Martineau.

R Core Team (2020). R: A Language and Environment for Statistical Computing [Computer Software Manual]. Vienna Austria: Retrieved from https://www.R-project.org.

Rauh, G. (1978). *Linguistische Beschreibung Deiktischer Komplexität in Narrativen Texten*. Tübingen: Narr Verlag.

Redeker, G. (1996). "Free Indirect Discourse in Newspaper Reports," in *Linguistics in the Netherlands 1996*. Editors C. Cremers and M. den Dikken (Amsterdam: Benjamins), 13, 221–232. doi:10.1075/avt.13.21red

Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27, 53–94.

Schlenker, P. (2004). Context of Thought and Context of Utterance: A Note on Free Indirect Discourse and the Historical Present. *Mind Lang.* 19 (3), 279–304. doi:10.1111/j.1468-0017.2004.00259.x

Schlenker, P. (2018). Gesture Projection and Cosuppositions. *Linguist Philos.* 41 (3), 295–365. doi:10.1007/s10988-017-9225-8

Schumacher, P. B., Dangl, M., and Uzun, E. (2016). *Thematic Role as Prominence Cue during Pronoun Resolution in German*. in *Empirical Perspectives on Anaphora Resolution*. Editors A. Holler and K. Suckow (Berlin: de Gruyter), 213–240.

Schumacher, P. B. L., Roberts, L., and Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua* 185, 25–41. doi:10.1016/j.lingua.2016.07.004

Sells, P. (1987). Aspects of Logophoricity. *Linguistic Inq.* 18, 445–479.

Sharvit, Y. (2008). The Puzzle of Free Indirect Discourse. *Linguist Philos.* 31 (3), 353–395. doi:10.1007/s10988-008-9039-9

Stec, K. (2016). Visible Quotation. PhD thesis. University of Groningen.

Stec, K. (2012). Meaningful Shifts. *Gest* 12 (3), 327–360. doi:10.1075/gest.12.3.03ste

Stokke, A. (2020). Free Indirect Discourse in Non-fiction. *Frontier Commun.* doi:10.3389/fcomm.2020.606616

Stokke, A. (2013). Protagonist Projection. *Mind Lang.* 28 (2), 204–232. doi:10.1111/mila.12016

Sundaresan, S. (2012). Context and (Co)reference in the Syntax and its Interfaces. PhD thesis. University of Stuttgart. doi:10.1109/infcom.2012.6195464

Yashima, Y. (2015). Antilogophoricity: In Conspiracy with the Binding Theory. PhD thesis. University of California at Los Angeles (UCLA).

Zeman, S. (2019). "Wer Spricht? Disambiguierungsfaktoren bei der Perspektivensetzung im Narrativen Diskurs," in *Rede- und Gedankenwiedergabe in narrativen Strukturen—Ambiguitäten und Varianz, Linguistische Berichte, Sonderheft*. Editors S. Engelberg, C. Fortmann, and I. Rapp (Hamburg: Helmut Buske Verlag), 27, 221–251.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.