**REVIEW ARTICLE**

WILEY

# Accounts of perspective taking in narrative

**Stefan Hinterwimmer** 🔟

Department of German studies,
University of Hamburg, Hamburg,
Germany

**Correspondence**
Stefan Hinterwimmer.
Email: Stefan.hinterwimmer@uni-hamburg.de

**Abstract**

This paper gives an overview over two different kinds of protagonists' perspective taking in narrative texts, *Free Indirect Discourse (FID)* and *Protagonist Projection (PP)/Viewpoint Shifting (VS)*, and the most important analyses of these phenomena that have been proposed within the framework of formal semantics and pragmatics. While FID is a special form of reporting self-reflexively conscious thoughts and utterances which in contrast to indirect and direct discourse is not overtly marked as such, PP/VS renders the content of protagonists' perceptions and beliefs. The paper discusses empirical differences between these two kinds of protagonists' perspective taking with respect to syntactic embeddability and the licencing of deictic expressions and considers various analytical options to capture these differences.

## 1 | INTRODUCTION AND OVERVIEW

Perspective taking is ubiquitous in natural language. Many descriptions involve a spatial perspective, that is, at least an implicit local center with respect to which other objects are located in space. Likewise, whenever a tense is chosen, this automatically requires at least an implicit temporal center with respect to which the reported situations can be located in the past, in the future or as co-temporal. In addition to temporal and spatial expressions, there are various other linguistic expressions that can only be interpreted relative to a perspective. This includes evaluative terms such as *beautiful* or *stupid*, expressives like *damn* or *fuck*, epithets like

*the idiot* and epistemic modals such as *must* or *might*: There always has to be someone who considers something beautiful or someone an idiot, who is in a heightened emotional state or with respect to whose belief state something is necessary or possible. In oral conversation, the default is for the respective speaker to be the perspective taker since speakers′ utterances are usually interpreted as expressing their thoughts or perceptions. Consequently, the speaker′s position in space at the utterance time is taken to be the region in space with respect to which spatial expressions are interpreted, the utterance time itself is the time with respect to which verbal tense marking is interpreted and it is the speaker′s taste and emotional or belief state with respect to which evaluative expressions, expressives, epithets and epistemic modals are interpreted (see Harris, 2012, 2021; Harris & Potts, 2009).

Sentences cannot only be understood as expressing the speaker′s thoughts or perceptions, however, but also as reporting thoughts, utterances or perceptions of other individuals. This can be indicated explicitly, as in (1a-c), where an event of saying (1a), thinking (1b) or perceiving (1c) is introduced by the respective matrix verb and where the complement of that verb renders the content of the respective event, whose agent or experiencer is the matrix subject.

(1) a. Jill said that she likes cats.
   b. Kamal thought that he had left his keys at home.
   c. Samira saw Bill leave the house without a jacket.

Note that the complements of the matrix verbs in (1a) and (1b) are finite clauses, while in (1c) the complement of the matrix verb consists of an infinite verb and its arguments and adjuncts. It is well known that not only in English, but in many other languages, the contents of speech and thought reports have to be rendered either by finite clauses or by DPs referring to abstract entities with content such as *claim*, *rumour* etc., which can take finite clauses providing the respective content as their complement (see Moulton, 2009, 2015 for detailed discussion). The contents of perception reports, in contrast, are typically rendered by constituents that can either be analysed as bare VPs containing a base generated subject, as infinite clauses or as NPs (see, e.g. Akmajian, 1977, Higginbotham, 1983; Safir, 1993 for discussion). This syntactic contrast between speech and thought reports, on the one hand, and perception reports, on the other, is quite plausibly reflective of a contrast in semantics: In case of the former two, the complement of the matrix verb is standardly analysed as denoting the set of possible worlds where the respective proposition is true, and the respective attitude verb expresses universal quantification over that set (Hintikka, 1969). The sentences in (1a) and (1b) are therefore interpreted as paraphrased in (2a) and (2b), respectively. The content of the complement of a perception verb, in contrast, is standardly analysed as a predicate characterising an event which functions as the theme of the event introduced by the perception verb (see Higginbotham, 1983, Barwise & Perry, 1983; Landman, 2000 for detailed discussion). The sentence in (1c) is therefore interpreted as paraphrased in (2c).

(2) a. There is a saying event *e* that is temporally located before the utterance time and whose agent is Mary and in all worlds that are compatible with the content of *e*, there is a state of Mary liking cats
   b. There is a thinking event *e* that is temporally located before the utterance time and whose experiencer is Kamal and in all worlds that are compatible with the content of *e* there is an event *e'* of Kamal leaving his keys at home that is temporally located before the time of *e*.
   c. There is a seeing event *e* that is temporally located before the utterance time and whose experiencer is Mary and there is an event *e'* of Bill leaving the house without his jacket and *e'* is the theme of *e*.

Concerning the contents of speech and thought events, they cannot only be rendered via *indirect discourse* (ID), as in (1a) and (1b), but also via *direct discourse* (DD), as in (3a) and (3b). Both (3a) and (3b) contain the first-person pronoun *I*, which belongs to the class of deictic expressions: The interpretation of deictic expressions directly depends on the context of utterance (Kaplan, 1989). Consequently, in each context of utterance *C* the first-person pronoun *I* denotes the speaker or thinker in *C*—from now on we will use the term *author of C*, which encompasses both roles—, the second person-pronoun *you* denotes the addressee in *C*, the temporal deictic expression *tomorrow* denotes the day following the day containing the time of *C*, etc. Crucially, in DD all deictic expressions are interpreted not with respect to the context of the reporting utterance, but with respect to the reported utterance or thought. Consequently, *I* in (3a) is interpreted as referring to Jill rather than the speaker of the **reporting utterance**, and likewise for (3b).

(3) a. Jill said: 'I like cats'.
  b. Kamal thought: 'I have left my keys at home'.

For ID, this is different, as becomes obvious by comparing (3a) with (1a) and (3b) with (1b): In (1a) and (1b), the pronoun referring to the respective speaker or thinker is a third-person pronoun, and if it is replaced by a first-person-pronoun, as in (4a) and (4b), the interpretation changes: (4a) can only be understood as reporting that Jill said that the author of the reporting utterance likes cats, and likewise for (4b). Note, however, that this difference between DD and ID regarding the interpretation of deictic expressions does not hold universally: As pointed out by Schlenker (2003) and much subsequent literature (see Deal, 2020 for a recent overview), first-person pronouns and other deictic expressions can be interpreted with respect to the context of the reported utterance in ID in many non-European languages. Even for languages such as English and German, at least temporal deictic expression such as *tomorrow* can be interpreted with respect to the reported utterance in ID for some speakers (see Plank, 1986; Anderson, 2020 for discussion).

(4) a. Jill said that I like cats.
  b. Kamal said that I had left my keys at home.

Regarding the behaviour of non-deictic perspective-dependent expressions such as evaluative expressions, expressives, epithets etc. in ID, there is a tendency for them to be interpreted with respect to the perspective of the individual whose thought or utterance is reported. In principle, however, they can be interpreted with respect to the speaker's or thinker's perspective as well. Consequently, the expressive *damn* in (5) is most likely understood as conveying Mary's perspective regarding Raoul's dog, although it can in principle also be understood as expressing the perspective of the speaker of the reporting utterance.

(5) Mary said that she would look after Raoul's damn dog over the weekend.

So far we have seen that in order for sentences to be understood as expressing the perspective of an individual different from the speaker, such a shift has to be indicated explicitly by the presence of a verb of saying, thinking, perceiving etc. This only holds outside of narrative texts, however. In narrative texts, a sentence or text segment can be interpreted as a thought, utterance or perception of a protagonist without having been introduced as such by the presence of an overt verb of saying, thinking or perceiving. To see this, consider the second sentence

in (6a), which is most plausibly interpreted as a thought that Maria has while or directly after staring at Raoul angrily rather than as a statement by the narrator. Likewise, the epithet *the idiot* is most naturally interpreted as expressing Maria's rather than the narrator's perspective.

(6) a-b. Maria stared at Raoul angrily. Today was her/*my birthday and the idiot had not even bought flowers.

Concerning the deictic expressions contained in (6a), they do not behave uniformly: On the one hand, the temporal adverbial *today* is interpreted with respect to Maria's perspective, that is, as referring to the day on which she has the thought that the sentence reports. On the other hand, only the third-person pronoun *her* in (6a) can be interpreted as referring to Maria, just as in the instances of ID in (1a) and (1b). For the first-person possessive pronoun *my* in the minimal variant in (6b), in contrast, such an interpretation is just as unavailable as the corresponding interpretation for the first-person pronouns in (4a) and (4b). Rather, in order to arrive at a sensible interpretation of (6b), the reader has to accommodate a so-called *homodiegetic narrator* as the referent of *my*. Such narrators correspond to fictional individuals that are part of the worlds depicted by the text and they stand in the same relation to the (propositions denoted by) the sentences that the respective text consists of as the speaker does to the sentences she utters in a conversation. Consequently, all perspectival expressions occurring in such a text can be assumed to be interpreted with respect to the narrator's perspective by default. Therefore, if a homodiegetic narrator is accommodated by the reader in order to arrive at a sensible interpretation of *my*, the second sentence in (6b) is no longer interpreted as a thought of Maria, but rather as a thought of the narrator, and both *tomorrow* and *the idiot* are interpreted with respect to narrator's perspective as well.

Out of the blue, however, it is very difficult to accommodate a homodiegetic narrator (see Altshuler & Maier, 2022 for relevant discussion) and the second sentence in (6b) is therefore more likely to be perceived as very awkward, due to the unexpected presence of the deictic pronoun *my*. The default case for the narrator of a fictional narrative text is to remain abstract: We do not get any information regarding the question of who tells the respective story and is accordingly responsible for the truth of the propositions the text consists of, when the story is told and where the individual telling the story is located at the time of telling. Nevertheless, it is standardly assumed in literary studies that fictional narrative texts are told by a narrator who takes responsibility for the truth of the propositions denoted by the sentences that comprise the text (see Martínez & Scheffel, 1999 for an overview). This is necessary in order to avoid the counterintuitive conclusion that the (vast majority of) the propositions denoted by the sentences contained in a fictional narrative are either lies told by the author or result from the author's having delusions.

Returning to the second sentence in (6a), which on its most plausible interpretation is interpreted as a thought of Maria rather than the narrator, it is worth noting that not only the possessive pronoun *her* but verbal tense marking, too, cannot be interpreted with respect to Maria's perspective. If it could, it would have to be present (perfect) rather than past (perfect): After all, Maria's thought concerns a state holding at the time at which she has that thought rather than a state temporally preceding that time. We will come back to the question of how the peculiar properties of sentences like the second sentence in (6a), which is an instance of *Free Indirect Discourse* (FID), can be accounted for in Section 2.1.

Consider next the sentence in (7), from Hinterwimmer (2017), and the one in (8), from Stokke (2013). On its most plausible interpretation, the main clause in (7) does not report an

event that is happening in the worlds compatible with the story (Lewis, 1978), that is, there is presumably no earthquake in those worlds. Rather, it reports a momentary illusion that Mary has as a consequence of her sense of balance being disturbed by the preceding boat trip. In (8), the second conjunct directly contradicts the first one, making it clear that the proposition denoted by that clause is not true in the worlds compatible with the story. Rather, it reports a false belief that the referent of the personal pronoun *her* has.

(7) When Mary stepped out of the boat, the ground was shaking beneath her feet for a couple of seconds.

(8) He gave her a ring studded with diamonds, but they turned out to be glass.

In Section 2.3 we will see that there are strong arguments for analysing sentences such as (7) and (8) as instances of a form of protagonists' perspective-taking that is distinct from FID and we will discuss three different analyses of this phenomenon, which has been dubbed *Viewpoint Shifting (VS)* by Hinterwimmer (2017) and *Protagonist Projection (PP)* by Stokke (2013, 2021) and Abrusán (2021b).

The main goal of this paper is to provide an overview over the various analyses that have been proposed for the two forms of protagonists' perspective-taking exemplified by (6a), on the one hand, and (7) and (8), on the other, and to consider some novel empirical data are puzzling for all existing accounts. The remainder of the paper is structured as follows: In Section 2.1 we will discuss the two most influential strands of analyses of FID that have been proposed in the literature and in Section 2.2 we will discuss the novel empirical data. In Section 2.3 we will discuss the three different analyses that have been proposed for the form of perspective-taking exemplified by the sentences in (7) and (8). Section 3 concludes.

## 2 | ANALYSES OF PERSPECTIVE-TAKING IN NARRATION

### 2.1 | Analyses of free indirect discourse

As we have already seen in Section 1, FID is a special form of thought (and also speech, as we will see below) representation which, in contrast to DD and ID, is not overtly marked as such, however: There is neither an embedding clause containing a propositional attitude verb such as *think* or *say* nor a preposed or postposed clause stating whose thought or utterance is rendered. Consequently, the reader has to infer that a sentence renders a thought or utterance of some contextually prominent protagonist (see Hinterwimmer, 2019; Abrusán, 2021a for discussion of the factors that make protagonists prominent as perspective takers) on the basis of the content in combination with indirect cues. Such cues are, for example, the presence of interjections such as *ouch* or *oops*, of evaluative expressions that cannot plausibly be interpreted as expressing the narrator's perspective and of deictic expressions that cannot plausibly be interpreted with respect to the narrator's context. These cues become particularly strong indicators that a sentence or text segment is to be interpreted as FID in cases where the narrator remains entirely abstract and is thus neither available as perspective taker nor as author of a context containing a clear and identifiable local or temporal parameter.

Regarding deictic expressions, their non-uniform behaviour in FID has already been mentioned in Section 1: While all other perspective-dependent expressions are interpreted with respect to the perspective of the protagonist whose thought or utterance is rendered, including local deictic adverbials such as *here* and *over there* and temporal deictic adverbials such as

*tomorrow* and *yesterday*, pronouns and tenses cannot be interpreted this way. Rather, they have to be interpreted with respect to the narrator's context. This holds in spite of the fact that this context is often entirely abstract since nothing whatsoever is known about the narrator, the time at which they tell the respective story and their location at that time—apart from the fact that the narrator is distinct from all the individuals occurring in the story since they are all referred to by third-rather than first-person pronouns (see Eckardt, 2015 for discussion) and that the temporal parameter of the narrator's context is located at a time later than the events reported in the story (if it is in past tense, that is). The non-uniform behaviour of deictic expressions is arguably the feature of FID that has received most attention from linguists working in the framework of truth conditional semantics, and the two most influential strands of analysis that have been developed in this tradition—double-context and mixed quotation accounts—mainly focus on accounting for it.

### 2.1.1 | Double-context analyses

While differing in many non-trivial details regarding the concrete technical implementation, the analyses by Schlenker (2004), Sharvit (2008) and Eckardt (2014) (see also Doron, 1991; Rauh, 1978 for earlier proposals based on similar ideas) share the same basic idea: Utterances cannot only be interpreted with respect to a single context, but with respect to two different contexts at the same time. The term *context* is understood in broadly Kaplanian (see Kaplan, 1989) terms, that is, as a tuple of parameters consisting of the author of the context, a time and a location, an, optionally, one or more addressee(s). For Schlenker (2004), all utterances are interpreted not just with respect to a single context, but with respect to two contexts: The context of utterance and the context of thought. In ordinary conversation, the two contexts coincide since the respective speaker utters their own thoughts. Consequently, the author of both contexts is the speaker, the temporal and local parameter of both is the utterance time and the location of the speaker at the utterance time, respectively, and the world parameter of both is the world where both the thought and the utterance occurs (setting aside the optional addressee parameter, which is irrelevant for our purposes). Likewise, in narration the two contexts coincide as long as the respective sentences are interpreted as expressing the narrator's perspective. Consequently, the author of both contexts is the narrator, and the other parameters are set accordingly, although, as already said, not only the narrator's identity, but also the time at which they are telling the story and their location at that time are often (almost; see above) completely unspecified. In FID, in contrast, the two contexts differ: The context of utterance still corresponds to the context of the narrator, but the context of thought is now the context of the protagonist whose thought or utterance is rendered, that is, the respective protagonist is the author of the context of thought and its temporal and local parameters are the time at which the thought or utterance occurs and the location of the protagonist at that time, respectively, where all these parameters are determined on the basis of information provided by the preceding narration.

Schlenker (2004) now assumes that all context-sensitive expressions with the exception of pronouns and tenses are lexically specified as being interpreted with respect to the context of thought: The temporal adverb *tomorrow*, for example, takes the context of thought as argument and maps it to the day following the day containing its temporal parameter. Pronouns and tenses (i.e. the tense markings of finite verbs), in contrast, are lexically specified as being interpreted with respect to the context of utterance. The first person pronoun *I*, for instance,

takes the context of utterance as argument and maps it to its author, the third person pronoun *he* maps it to a salient male individual distinct from the author of the context of thought as well as from the addressee(s), and the past tense morpheme maps it to a time preceding its temporal parameter. In ordinary conversation as well as in standard narration, it does not make any difference that pronouns and tenses are interpreted with respect to the context of utterance, while all other deictic expressions are interpreted with respect to the context of thought, since the two contexts coincide anyway. In FID, however, where the two contexts are actually different, it automatically leads to the characteristic inconsistent interpretation of deictic expressions.

The accounts of Sharvit (2008) and Eckardt (2014) are based on the same underlying idea. Concerning Sharvit (2008), the main difference to the account of Schlenker (2004) is that the contexts with respect to which all deictic expressions except pronouns and tenses are interpreted are introduced by a covert operator: a silent counterpart of *think* which quantifies over contexts instead of possible worlds (I will come back to the issue of how the partial context shift characteristic for FID is related to syntax in Section 2.2). In the absence of that operator, utterances are interpreted with respect to a single context of utterance. Concerning Eckardt (2014), the first main difference to the account of Schlenker (2004) is that she does not assume deictic expressions to be functions from contexts to individuals, times, locations and worlds, respectively. Rather, in her system contexts themselves are functions taking the respective expressions as arguments and mapping them to individuals, times, locations and worlds, respectively. The reasons for this technical change are not crucial for the purposes of this paper, so we will not discuss them for reasons of space. What is more important for our purposes, however, is the following point: In contrast to the other double context analyses discussed so far, Eckardt (2014) is quite explicit about the way in which the propositions denoted by sentences interpreted as FID are combined with the propositions denoted by the other sentences the respective narrative consists of. She assumes that the set of worlds in which all the propositions denoted by the sentences comprising the narrative up to this point are not intersected directly with the proposition denoted by the sentence interpreted as FID. Rather, the former set of worlds is intersected with the set of worlds in which there is a temporally overlapping thinking or speaking event whose agent or experiencer is the respective protagonist and whose content is the proposition interpreted as FID. The two sentences in (6a), repeated here as (9a), are thus interpreted as (roughly) paraphrased in (9b).

(9) a. Maria stared at Raoul angrily. Today was her birthday and the idiot had not even bought flowers.
b. In all worlds $w$ that are compatible with the story so far there is an event $e$ of Maria staring at Raoul angrily and $e$ took place before the time of $C$, and there is a temporally overlapping thinking event $e'$ and the experiencer of $e'$ is Maria and in all worlds $w'$ that are compatible with the content of $e'$, there is a state of Maria having birthday that is located at the day containing the time of $c$ and before the time of $C$ and there is no event $e''$ of the most prominent individual that Maria considers to be an idiot at the time of $c$ buying flowers such that $e''$ took place before the reference time of the ongoing story, and the reference time is located before the time of $C$, where $C$ is the narrator's context and $c$ the protagonist's (i.e. Maria's) context.

In the following section, we will consider an alternative account which does not involve a second context, but rather assumes FID to be special kind of mixed quotation.

## 2.1.2 | The mixed-quotation account

Maier (2015, 2017; see also Dirscherl & Pafel, 2015) proposes an account of FID which is based on the idea that FID is a special, highly conventionalised form of mixed quotation in which pronouns and tenses (i.e. tense markings on verbs) are systematically unquoted. In mixed quotation, which is often found in journalistic texts, an utterance is partially quoted, while other parts receive their standard interpretation. In journalistic texts, the quoted parts are usually typographically marked and thus set apart from the unquoted parts by quotation marks, italics, etc. According to Maier (2015, 2017), the absence of any typographic marking in FID is due to the high degree of conventionalisation. In his view, the high degree of conventionalisation is also responsible for the systematic unquotation of pronouns and tenses in FID, which is found to a lesser degree in journalistic texts as well (see Maier, 2017 for detailed discussion). Simplifying somewhat, since the tense markings cannot be separated from the verbal stems, the second sentence in (6a)/(9a) corresponds to the sentence in (10), in which mixed quotation is explicitly marked typographically.

(10)     'Tomorrow' was her 'birthday and the idiot' had 'not even bought flowers'.

Concerning the interpretation of sentences involving mixed quotation, Maier (2015, 2017) assumes that events of speaking or thinking consist of concatenated subevents that have a content as well as a form and that in mixed quotation the unquoted parts correspond to subevents whose content is given, while the quoted parts correspond to subevents whose form is given. Consequently, in order to interpret a sentence in FID, a speaking or thinking event has to be accommodated whose agent or experiencer is the respective contextually prominent protagonist, and the forms of the subevents corresponding to all the linguistic material except pronouns and tenses are concatenated with the forms of the subevents corresponding to the pronouns and tenses, of which not the form, but only the content is given. The sentence in (10) is thus interpreted as paraphrased in (11).

(11)     There is a thinking event $e$ whose experiencer is Maria and there are subevents $e_1$ and $e_2$ of $e$, and the form of $e$ is *tomorrow* concatenated with the form of $e_1$ concatenated with *birthday and the idiot* concatenated with the form of $e_2$ concatenated with *not even bought flowers*, and the content of $e_1$ is the denotation of *was her* and the content of $e_2$ is the denotation of *had*.

Both double-context and mixed-quotation analyses capture the basic interpretative properties of FID in a formally precise way, and both accounts need to be combined with a pragmatic account of the conditions which make protagonists available as implicit perspective takers (see Hinterwimmer, 2019; Abrusán, 2021a for discussion). Concerning the question of which of the two types of analyses is to be preferred, Maier (2015, 2017) provides the following argument in favour of the mixed-quotation analysis: There are instances of FID where the utterances of a protagonist are given in their non-standard dialect, while the surrounding text is given in standard language. Such cases are easily accounted for by a mixed-quotation analysis, but it is completely unclear how they are to be captured by double-context analyses. As pointed out by Eckardt (2014), however, such cases are confined to the representation of utterances, not thoughts. The second sentence in (12), for instance, which is rendered in Bavarian dialect, while the first sentence is in standard German, can only be interpreted as (a rendition of) an utterance of Sepp, not as a thought—to the extent that it is acceptable at all, given the absence of contextual cues for the introduction of an utterance situation.

(12)     Sepp stand wütend auf. Da Bäda woa a Saudepp!
         *Sepp stood up angrily. Peter was a fuckin' idiot*!

In Section 2.2 we will see that there are additional empirical differences between speech and thought representation via FID.

## 2.2 | Empirical differences between speech and thought representation via FID

In Section 2.1.1 I have introduced Sharvit's (2008) double context analysis of FID, according to which the contexts with respect to which all context-sensitive expressions with the exception of pronouns and tenses are interpreted are introduced by a covert operator. Now, there are strong arguments that FID is incompatible with syntactic embedding. First, as already pointed out by Banfield (1982), sentences expressing speech acts such as exclamations and optatives, which cannot be embedded, can be interpreted as FID. Second, Hinterwimmer (2017) observes that the shift from standard narration to FID cannot occur sentence-internally, but only at the root level. To see this, consider the contrast between (13a) and (13b), on the one hand, and (13c), on the other, from Hinterwimmer (2017).

(13)     a. On her way home, Mary heard a song by Kendrick Lamar that she liked
         on the radio. She would buy his new album tomorrow.
         b. When Mary heard a song by Kendrick Lamar that she liked on the radio
         on her way home, she thought: "I will buy his new album tomorrow".
         c. * When Mary heard a song by Kendrick Lamar that she liked on the radio
         on her way home, she would buy his new album tomorrow.

While the second sentence in (13a) can easily be interpreted as a thought that Mary has while hearing a song by Kendrick Lamar that she likes on the radio, a parallel interpretation is completely excluded for the main clause in (13c), which is hard to make sense of at all. Note that the only difference between (13a) and (13c) is that in the latter the information that is given in the form of an independent main clause in the former is given in the form of a left-adjoined *when*-clause. Crucially, the relation of temporal overlap that is enforced by *when* in the case of (13c) holds in (13a) as well, that is, the thinking event whose content is given by the second sentence in (13a) is understood as temporally overlapping with the hearing event introduced by the first sentence. Finally, as shown by the acceptability of (13b), there is of course no inherent problem with the embedding of thinking events, as long as those events are introduced by overt linguistic material. What seems to be excluded, however, is the sentence-internal accommodation of implicit thinking or speaking events and/or the partial context shift that is required for FID.

Now, if FID comes about via the introduction of an additional context, the unavailability of an FID-interpretation of the main clause in (13c) might well be due to the impossibility to introduce such a context sentence-internally, that is, it is conceivable that contexts can only be introduced at the root level, which is the level where sentences can be interpreted as speech acts. If FID is just a special form of mixed quotation, in contrast, the unavailability of an FID-interpretation of the main clause in (13c) could plausibly be assumed to be due to the impossibility to accommodate the speech or thought acts to be partially quoted sentence-internally.

The latter claim can directly be tested, since it should also apply to instances of mixed quotation where the quoted parts are typographically marked as such. With this in mind, consider the following example, which is fully acceptable.

(14)    When Claire told Charles that she was the new chief editor of *Linguistics or Philosophy?*, she was suddenly 'the greatest philosopher not only in this room, but in the entire world' and he would 'certainly lend [her his] new Ferrari tonight'.

Now, in order to arrive at a coherent interpretation, the speech event whose content is partially quoted in (14) has to be accommodated sentence-internally, that is, at the level of the main clause: It is only after having been informed about Claire's new position that Charles claims her to be the greatest philosopher.

In light of the unavailability of an FID-interpretation for the main clause in (13c), the acceptability of (14) could at first sight be taken as an argument against an analysis of FID as mixed quotation. This would be premature, however, since the variant of (14) in (15), which does not contain any typographically marked elements, is just as fine and receives the same interpretation as (14).

(15)    When Claire told Charles that she was the new chief editor of *Linguistics or Philosophy?*, she was suddenly the greatest philosopher not only in this room, but in the entire world and he would certainly lend her his new Ferrari tonight.

In other words, the main clause in (15) is interpretable as FID, contrary to what we have seen in (13c) above. Note that the main clause (15) contains two deictic expressions, namely *this room* and *tonight*, which are both most naturally interpreted with respect to the context in which Charles' utterance occurred.

Now, an obvious difference between (13c) and (15) is that while in the former a thought is rendered as FID, in the latter it is an utterance. Further evidence for the assumption that the contrast between speech and thought representation is the crucial point is provided by the variant of (13c) given in (16), which, in contrast to (13c), allows the main clause to be interpreted as FID. Crucially, however, this is only possible if the main clause is interpreted as rendering an utterance, not a thought of Mary. This is quite plausible in the context provided by the *when*-clause in (16), but not in the context provided by the one in (13c).

(16)    When Charles told Mary that the song by Kendrick Lamar that they heard on the radio was from a Pulitzer Prize winner, it suddenly was the greatest song she had heard this year and she would certainly buy his new album tomorrow

In light of the difference between speech and thought representation via FID regarding renditions of non-standard dialects that we have seen in Section 2.1, one option that comes to mind would be to assume that mixed quotation is the right analysis for speech renditions via FID, while thought renditions come about via the introduction of a protagonist's context. If one additionally assumes that protagonists' contexts can only be introduced at the root level, the contrast between (13c) and (15) is automatically accounted for. Such an approach is rather unattractive, however, since two quite different mechanisms have to be invoked for otherwise very similar phenomena, which raises more questions than it answers. If one wants to maintain a unified approach of FID, either via mixed quotation or via an additional protagonist's context, in contrast, the following option comes to mind: One could assume that the accommodation of

speech events (which are either partially quoted or which introduce protagonists' contexts) is possible at any level. The accommodation of thought events, in contrast, (which are likewise either partially quoted or which introduce protagonists' contexts) could be assumed to be restricted to the root level exclusively. There are two problems with such an approach, however: First, in the absence of any independent motivations for making such a distinction, it is completely ad hoc. Second, it is in conflict with the data on VS/PP to be discussed in detail in Section 2.3, which involve thought events that can be embedded. Therefore, the contrast between speech and thought reports via FID regarding embeddability remains an open question that has to be addressed by future research.

## 2.3 | Analyses of viewpoint shifting/protagonist projection

In Section 1 I have introduced the sentences in (7) (from Hinterwimmer (2017)) and (8) (from Stokke (2013)), repeated here as (17) and (18), respectively, as a form of protagonists' perspective taking in narratives which is different from FID in the following respect: It does not render the content of a conscious thought. Rather, it reports the content of a perception in a way that is compatible with the belief state of the respective protagonist at the time of perceiving.

(17)    When Mary stepped out of the boat, the ground was shaking beneath her feet for a couple of seconds.

(18)    He gave her a ring studded with diamonds, but they turned out to be glass.

Before we will turn to three different analyses that have been proposed for this phenomenon, it is worth considering (17) in light of the discussion in Section 2.2, where we have seen that the shift from standard narration to FID cannot happen sentence-internally if a thought is rendered. Now, in (17) the shift from standard narration to Mary's perspective clearly happens sentence-internally: The *when*-clause is a neutral description of an event which happens in the worlds compatible with the story, while the main clause describes an event that (on the most natural interpretation) only happens in the worlds that are compatible with Mary's beliefs at the time of her sense of balance being temporarily disturbed by a preceding boat trip. The availability of this interpretation to the main clause in (17) therefore provides an argument that the phenomenon under discussion is distinct from FID. Consider the contrasts between (19a), (19b) and (19c) from Hinterwimmer (2020) in this light.

(19)    a. Mary looked up from her whiskey glass. Gosh, the counter over there was moving up and down continuously.
b. ??When Mary looked up from her whiskey glass, the counter over there was moving up and down continuously.
c. When Mary looked up from her whiskey glass, the counter on the other side of the room was moving up and down continuously.

The second sentence in (19a) can easily be interpreted as a conscious thought of Mary rendered as FID. For (19b), in contrast, a parallel interpretation of the main clause as FID is much harder, similarly to the case of (13a) versus (13c). At the same time, the deictic adverbial *over there* can only sensibly interpreted with respect to Mary's context, that is, an alternative interpretation is not available. Finally, the main clause in (19c) can easily be interpreted as rendering the content of a perception of Mary in a way that is compatible with her belief state at the time of perceiving. Crucially, however, this reading is only available because the deictic

adverbial *over there* has been replaced by the non-deictic one *on the other side of the room*. The variant in (19b) with the deictic adverbial does not allow such an interpretation, but only one as FID, which is blocked for syntactic reasons, however. The contrast between (19b) and (19c) therefore not only shows that the form of perspective-taking discussed in this section is compatible with syntactic embedding, but also that it does not allow a shifted interpretation of deictic elements.

## 2.3.1 | The operator account

In Hinterwimmer (2017) the form of protagonist perspective-taking exemplified by (7)/(17), (8)/(18) and (19c)[1] is dubbed *VS* and an analysis is proposed according to which VS comes about via the insertion of a covert operator. This operator, which is assumed to be optionally inserted at the level of finite clauses, introduces a perception event, a free variable ranging over individuals and a free variable ranging over perception predicates and it takes the content of the clause to which it is adjoined as argument. The free variable ranging over individuals is resolved to a contextually prominent discourse referent and the free variable ranging over perception predicates to a perception predicate that is plausible in the context, that is, a predicate such as *see*, *feel*, *hear* etc. The clause to which the operator has been adjoined is then interpreted as characterising an event which takes place in all the worlds that are compatible with the belief state of the protagonist who is the experiencer of the perception event introduced by the operator. Assuming that the free individual variable is resolved to Mary and the free predicate variable to *see(ing)*, the main clause in (19c), for example, is then interpreted as paraphrased in (20a), and the entire sentence as paraphrased in (20b).

(20)   a. There is a seeing event $e_1$ whose experiencer is Mary and in all worlds that are compatible with Mary's belief state at the time of $e_1$, there is an event $e_2$ of the counter on the other side of the room moving up and down continuously.
b. There is a past event $e$ of Mary looking up from her whiskey glass and there is a temporally overlapping seeing event $e_1$ whose experiencer is Mary and in all worlds that are compatible with Mary's belief state at the time of $e_1$, there is an event $e_2$ of the counter on the other side of the room moving up and down continuously.

Since the mechanism involved in VS is fundamentally different from the one involved in FID, the operator approach can account for the differences between thought reports via FID and VS regarding syntactic embeddability and the differences between FID (in general) and VS/PP regarding the interpretation of deictic expressions: The former involves the introduction of an additional protagonist's context with respect to which all context-sensitive expressions except pronouns and tenses are interpreted, and this is only possible at the root level according to Hinterwimmer (2017). The latter, in contrast, does not involve the introduction of an additional context with respect to which any deictic expressions could be interpreted. Rather, it involves the insertion of a covert operator quantifying over possible worlds, which is possible on the level of finite clauses. The operator approach does not have an answer to the question of why speech reports via FID can be embedded, too, however. The second downside of the operator approach is that it has to stipulate a covert operator in addition to the mechanism that accounts for FID.

### 2.3.2 | The double-context account

Abrusán (2021b) dubs the phenomenon exemplified by (7)/(17), (8)/(18) and (19c) *PP*, following Stokke (2013). She proposes an analysis which is a rather minimal extension of Eckardt's (2014) account of FID, that is, she assumes PP/VS to involve the introduction of an additional protagonist's context, too. The only difference to FID is that the additional context in PP is not introduced by an implicit speaking or thinking event, but rather by a perceiving or believing event whose experiencer is some contextually prominent protagonist. Consequently, propositions that are evaluated with respect to such contexts are then not interpreted as the contents of the respective speaking or thinking events, as in FID, but rather as the contents of the respective perceiving or believing events. The main clause in (19c) thus receives an interpretation Abrusán's (2021b) account that is essentially equivalent to the one paraphrased in (20b), but it comes about in a way that only minimally differs from the one bringing about FID. In order to account for the observation that deictic expressions generally[2] do not receive a shifted interpretation in PP/VS (as shown by the contrast between (17b) and (17c)), Abrusán (2021b) assumes them to come with a special presupposition. This presupposition requires the context with respect to which they are interpreted to be the context of an overt or covert (in the case of FID rendering thoughts) speech act.

Abrusán's (2021b) analysis is attractive insofar as it introduces virtually no additional machinery and successfully captures the interpretative properties of PP/VS. Just like the operator approach, however, it does not have an answer to the question of why thought reports via FID are restricted to root clauses, while both VS/PP and speech reports via FID can be introduced sentence-internally.

### 2.3.3 | The mixed-quotation account

Stokke (2021) proposes an analysis of PP/VS that is a rather minimal extension of Maier's (2015, 2017) mixed-quotation account of FID. In Stokke's view, the main difference between FID and PP/VS is that in the latter what is partially quoted is not the content of an actual (overt or covert) speech event, but rather the content of a possible speech event: something that the respective protagonist could have uttered given their beliefs and dispositions. The main clause in (19c) would thus be interpreted on Stokke's (2021) analysis as partially (i.e. with the exception of the tensed verb) quoting an utterance Mary could have made, given her beliefs and dispositions, at the moment at which she was looking up from her whiskey glass.

Just like the account of Abrusán (2021a, b), the analysis proposed by Stokke (2021) is attractive insofar as it requires almost no machinery beyond the one that has been proposed for analysing FID anyway. It faces the following challenges, though, the first of which it shares with the other two approaches discussed in this section: If both FID and PP/VS come about via mixed quotation, why is FID rendering thoughts restricted to root clauses, while PP/VS and FID rendering utterances may occur sentence-internally? Second, why can deictic expressions occurring in actual (overt or covert) utterances be partially quoted, while ones occurring in possible utterances cannot?

# 3 | CONCLUSION

In this paper I have discussed three different kinds of implicit protagonists' perspective taking in narrative texts: Thought reports via FID, speech reports via FID and perception reports via PP/ VS. I have discussed recent analyses of these phenomena that have been proposed within the framework of truth conditional semantics and formal pragmatics and have presented novel empirical evidence showing that FID rendering the content of utterances behaves differently from FID rendering the content of thoughts with respect to syntactic embeddability. Regarding the space of analytical options to account for the three kinds of perspective taking, on one end of the spectrum there is the option to invoke a different mechanism for each kind: Mixed quotation for speech reports via FID, the introduction of a protagonist's context in addition to the narrator's context for thought reports via FID, and a covert operator quantifying over worlds compatible with the belief state of a prominent protagonist at the time of a contextually provided perceiving eventuality. On the other end of the spectrum, there is the option to analyse all three kinds of perspective taking via the same mechanism—either via mixed quotation or via double contexts. While both unified approaches are in principle highly attractive in terms of theoretical economy, they face the challenge of accounting for the differences with respect to syntactic embeddability between speech representation via FID and VS/PP, on the one hand, and thought representation via FID, on the other, and the differences regarding the availability of shifted interpretations for deictic expressions between speech and thought representations via FID, on the one hand, and VS/PP, on the other. More research is required to see whether convincing accounts of these differences can be developed within the two uniform approaches, or whether a non-uniform analysis is ultimately preferable.

## CONFLICT OF INTEREST STATEMENT
The author has no competing interests to declare.

## ORCID
*Stefan Hinterwimmer* https://orcid.org/0000-0002-2808-8361

## ENDNOTES
[1] In Hinterwimmer (2017) examples like those in (18), which have been dubbed Protagonist Projection by Stokke (2013) are not discussed, but Abrusán (2021b) argues convincingly that *Viewpoint Shifting* and *Protagonist Projection* are the same phenomenon.

[2] But see Abrusán (2021b) for some exceptions, which are the same that can receive a shifted interpretation in indirect discourse as well and therefore arguably are not really context-sensitive in the sense of Kaplan (1989).

## REFERENCES
Abrusán, M. (2021a). Computing perspective shift in narrative texts. In E. Maier & A. Stokke (Eds.), *The language of fiction* (pp. 325–348). Oxford University Press.

Abrusán, M. (2021b). The spectrum of perspective shift: Free indirect discourse vs. protagonist projection. *Linguistics and Philosophy*, *44*(4), 839–873. (01/20/2022). https://doi.org/10.1007/s10988-020-09300-z

Akmajian, A. (1977). The complement structure of perception verbs in an autonomous syntax framework. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 427–460). Academic Press.

Altshuler, D., & Maier, E. (2022). Coping with imaginative resistance. *Journal of Semantics*, *39*(3), 523–549. https://doi.org/10.1093/jos/ffac007

Anderson, C. (2020). *Shifting the perspectivallLandscape: Methods for encoding, identifying, and selecting Perspectives*. PhD thesis. University of Massachusetts.

Banfield, A. (1982). *Unspeakable sentences: Narration and representation in the language of fiction*. Routledge.

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. MIT Press.

Deal, A. R. (2020). *A theory of indexical shift. Meaning, grammar, and crosslinguistic variation*. MIT Press.

Dirscherl, F., & Pafel, J. (2015). Die vier Arten der Rede-und Gedankendarstellung Zwischen Zitieren und Referieren. *Linguistische Berichte*, *241*, 3–47. https://doi.org/10.46771/2366077500241_1

Doron, E. (1991). Point of view as a factor of content. *Proceedings of Semantics and Linguistic Theory*, *1*, 51–64. https://doi.org/10.3765/salt.v10i0.2743

Eckardt, R. (2014). *The Semantics of free indirect discourse: How texts allow to mindread and eavesdrop*. Brill.

Eckardt, R. (2015). Speakers and narrators. In D. Birke & T. Köppe (Eds.), *Author and narrator: Transdisciplinary contributions to a narratological debate* (pp. 153–186). de Gruyter.

Harris, J. A. (2012). *Processing perspectives*. PhD thesis. University of Massachusetts.

Harris, J. A. (2021). Extended perspective shift and discourse economy in language processing. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.613357

Harris, J. A., & Potts, C. (2009). Perspective-shifting with appositives and expressives. *Linguistics and Philosophy*, *32*(6), 523–552. (01/20/2022). https://doi.org/10.1007/s10988-010-9070-5

Higginbotham, J. (1983). The logic of perceptual reports: An extensional situation semantics. *Journal of Philosophy*, *80*(2), 100–127. https://doi.org/10.2307/2026237

Hinterwimmer, S. (2017). Two kinds of perspective taking in narrative texts. *Proceedings of Semantics and Linguistic Theory*, *27*, 282–301. https://doi.org/10.3765/salt.v27i0.4153

Hinterwimmer, S. (2019). Prominent protagonists. *Journal of Pragmatics*, *154*, 79–91. https://doi.org/10.1016/j.pragma.2017.12.003

Hinterwimmer, S. (2020). Zum Zusammenspiel von Erzähler- und Protagonistenperspektive in den Brenner-Romanen von Wolf Haas. In S. Zeman (Ed.) *Zeitschrift für germanistische Linguistik, 48*, special issue *Narrativität als linguistische Kategorie* (pp. 529–561). https://www.degruyter.com/journal/key/zfgl/48/3/html

Hintikka, J. (1969). Semantics for propositional attitudes. In J. W. Davis, D. J. Hockney, & W. K. Wilson (Eds.), *Philosophical logic (synthese library 20)* (pp. 21–45). Springer.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 565–614). Oxford University Press.

Landman, F. (2000). *Events and plurality: The Jerusalem lectures (2000)*. Kluwer.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly*, *15*(1), 37–46.

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind & Language*, *30*(3), 345–373. https://doi.org/10.1111/mila.12083

Maier, E. (2017). The pragmatics of attraction. Explaining unquotation in direct and free indirect discourse. In P. Saka & M. Johnson (Eds.), *The semantics and pragmatics of quotation* (pp. 259–280). Springer.

Martínez, M., & Scheffel, M. (1999). *Einführung in die Erzähltheorie*. C. H. Beck.

Moulton, K. (2009). *Natural selection and the syntax of clausal complementation*. PhD thesis. University of Massachusetts.

Moulton, K. (2015). CPs: Copies and compositionality. *Linguistic Inquiry*, *46*(2), 305–342. https://doi.org/10.1162/ling_a_00183

Plank, F. (1986). Über den Personenwechsel und den anderer deiktischer Kategorien in der wiedergegebenen Rede. *Zeitschrift für Germanistische Linguistik*, *13*(3), 284–308. https://doi.org/10.1515/zfgl.1986.14.3.284

Rauh, G. (1978). *Linguistische Beschreibung deiktischer Komplexität in narrativen texten*. Narr.

Safir, K. (1993). Perception, selection, and structural economy. *Natural Language Semantics*, *2*(1), 47–70. https://doi.org/10.1007/bf01255431

Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy*, *26*(1), 29–120. https://doi.org/10.1023/A:1022225203544

Schlenker, P. (2004). Context of thought and context of utterance. A note on free indirect discourse and the historical present. *Mind & Language*, *19*(3), 279–304. https://doi.org/10.1111/j.1468-0017.2004.00259.x

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy*, *31*(3), 353–395. https://doi.org/10.1007/s10988-008-9039-9

Stokke, A. (2013). Protagonist projection. *Mind & Language*, *28*(2), 20–232. https://doi.org/10.1111/mila.12016

Stokke, A. (2021). Protagonist projection, character focus and mixed quotation. In E. Maier & A. Stokke (Eds.), *The Language of fiction* (pp. 379–404). Oxford Univ Press.

## AUTHOR BIOGRAPHY

**Stefan Hinterwimmer** research interests lie in the realms of semantics, pragmatics, the syntax-semantics interface, information structure and text linguistics. He has worked on adverbial quantification, free relative clauses, conditionals, topicality, specificity, pronoun resolution, the contrast between definite and demonstrative DPs and perspective taking. He received his PhD in 2006 from Humboldt-Universität zu Berlin. He is currently Professor at the Department of German studies at the University of Hamburg.