Taking the perspective of narrative characters: a mouse-tracking study on the processing of ambiguous referring expressions in narrative discourse

Jorrig Vogels[1], Sofia Bimpikou[1], Owen Kapelle[1], Emar Maier[1,2]

[1]Center for Language and Cognition Groningen (CLCG), Faculty of Arts, University of Groningen, Groningen, The Netherlands

[2]Department of Theoretical Philosophy, Faculty of Philosophy, University of Groningen, Groningen, The Netherlands

Correspondence:

Jorrig Vogels
Faculty of Arts
Semantics and Cognition
Oude Kijk in 't Jatstraat 26
9712 EK Groningen
The Netherlands
j.vogels@rug.nl

Sofia Bimpikou
Faculty of Arts
Semantics and Cognition
Oude Kijk in 't Jatstraat 26
9712 EK Groningen
The Netherlands
s.bimpikou@rug.nl

Owen Kapelle
Faculty of Arts
Semantics and Cognition
Oude Kijk in 't Jatstraat 26
9712 EK Groningen
The Netherlands
o.e.j.kapelle@student.rug.nl

Emar Maier
Faculty of Arts
Semantics and Cognition
Oude Kijk in 't Jatstraat 26
9712 EK Groningen
The Netherlands

Faculty of Philosophy
Theoretical Philosophy
Oude Boteringestraat 52
9712 GL Groningen
The Netherlands
e.maier@rug.nl

RUNNING HEAD: Taking the perspective of narrative characters

Taking the perspective of narrative characters: a mouse-tracking study on the processing of ambiguous referring expressions in narrative discourse

Abstract
An ongoing debate in the interpretation of referring expressions concerns the degree to which listeners make use of perspective information during referential processing. We aim to contribute to this debate by considering perspective shifting in narrative discourse. In a web-based mouse-tracking experiment in Dutch, we investigated whether listeners automatically shift to a narrative character's perspective when resolving ambiguous referring expressions, and if different linguistic perspective-shifting devices affect how and when listeners switch to another perspective. We compared perspective-neutral, direct, and free indirect discourse, manipulating what objects are visible to the character. Our results do not show a clear effect of the perspective shifting devices on participants' eventual choice of referent, but our online mouse-tracking data reveal processing differences that suggest that listeners are indeed sensitive to the conventional markers of perspective shift associated with direct and (to a lesser degree) free indirect discourse.

Keywords: perspective shifting, narrative, free indirect discourse, mouse tracking, reference resolution.

1. Introduction

Perspective taking plays a crucial role in language comprehension. In conversation, it has been shown that listeners take into account the speaker's point of view in order to resolve possibly ambiguous referring expressions (e.g. Hanna et al., 2003; Nadig & Sedivy, 2002). For instance, upon hearing "Pick up the big…", listeners are quick to identify the target object as the largest object of a size contrast (e.g. a big and a small duck), even when there is a second size contrast visible to the listener but not to the speaker (Heller et al., 2008). This finding suggests that during reference processing listeners reason that the adjective 'big' most likely relates to the size contrast that is actually available to the speaker.

Perspective taking is also fundamental when engaging in narrative discourse. In narratives, readers or listeners have to consider the perspective of narrative characters in order to distinguish what's actually happening in the story world from what the characters think, say, imagine, or see (e.g. Ferguson et al., 2015; Salem et al., 2017; Köder & Maier, 2018). The required perspective shifts may be encoded grammatically by report constructions like direct discourse or free indirect discourse. In direct discourse (DD), there is an explicit reporting clause, or frame, and all context-sensitive expressions (pronouns, tenses, expressives) refer to the reported speaker. Free indirect discourse (FID) is a mixed form of reporting in that it combines features from both direct and indirect discourse (Fludernik, 2003). Specifically, just as in indirect discourse (ID), pronouns and tenses in FID are interpreted with reference to the reporting (i.e. the narrator's) context, whereas everything else, e.g., exclamatives and other speaker-oriented constructions, is interpreted as quoted, i.e. with reference to the protagonist, as is the case in DD.

Little is known about whether and how this type of narrative perspective shifting affects referential processing (but see Köder et al., 2015, and Köder & Maier, 2018). In this paper, we therefore explore the time course of narrative perspective taking by tracking the resolution of perspective-ambiguous referring expressions during real-time processing of narratives. Specifically, we investigate to what degree listeners disambiguate referring expressions by taking the narrative character's perspective into account, and whether the use of perspective-shifting markers encourages listeners to switch to the character's perspective To this end, we conducted a Visual World Paradigm mouse-tracking study in

Dutch, in which participants listened to stories such as given in (1) below that contained references to objects in a visual scene (e.g. *het grote varken* 'the big pig'; see Figure 1). We manipulated both whether the objects were visually shared between listener and character (Shared, Non-shared) and the presence of perspective-shifting cues in the narratives (Perspective-neutral, Direct, or Free indirect discourse), resulting in a 2x3 within-item design. In the Shared condition (left side of Figure 1), the listener's visual perspective coincides with the visual perspective of the narrative character: the target object (here, the biggest pig) and all possible competitors (a medium-sized pig and a small pig) are visible to both the character and the listener. In the Non-shared condition (right side of Figure 1) the two perspectives are dissociated: an obstacle restricts which objects are visible to the character. Here, the best-fitting referent from the listener's perspective is visually inaccessible to the character. For instance, the expression 'the big pig' is ambiguous when the biggest pig in the listener's view is visually hidden from the character. For the character, the best-fitting referent is the biggest among the pigs that are accessible to her.

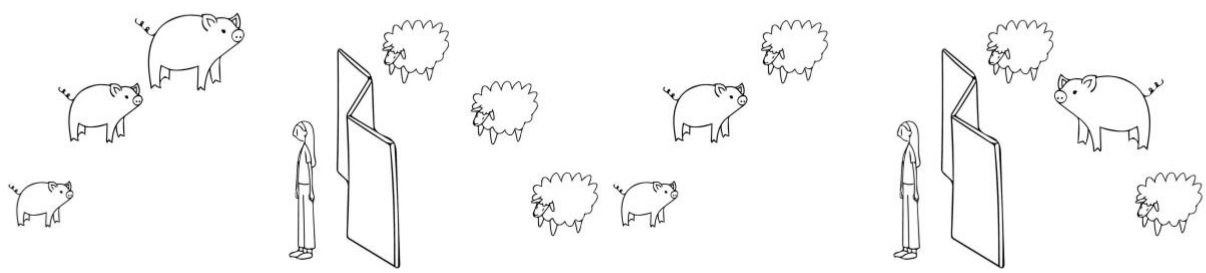

Figure 1. Example stimulus. Left: Shared condition; Right: Non-shared condition.

(1) *Laura stond te dromen in de schuur. De varkens scharrelden onrustig in het rond.*
'Laura stood dreamily in the shed. The pigs were restlessly pottering about.'
   a. *Het grote varken werd erg chagrijnig.* **(Neutral Discourse)**
      'The big pig got really annoyed.'
   b. *Ze dacht: "Oh oh, het grote varken wordt erg chagrijnig!"* **(Direct Discourse)**
      'She thought: "Uh oh, the big pig is getting really annoyed!"'
   c. *Oh oh, het grote varken werd erg chagrijnig!* **(Free Indirect Discourse)**
      'Uh oh, the big pig was getting really annoyed!'
Het was niet zo slim om er alleen naar toe te gaan.
'It wasn't very wise to go there alone.'

We use the perspective-neutral discourse (ND) condition (1a) as a baseline to see to what extent listeners move their mouse towards the character-accessible referent in the absence of perspective-shifting cues. In the direct discourse (DD) condition (1b), we include an initial overt frame ('she thought'), an exaggerated mimetic intonation of the quoted sentence, and an exclamative at the start of the quoted phrase (e.g. 'uh oh') to explicitly mark shifting to the narrative character's perspective. In the free indirect discourse (FID) condition, we relied exclusively on a vivid, mimetic intonation and the initial exclamative to mark the perspective shift. In order to investigate listeners' time course of reference processing, we measure their mouse trajectories from the offset of the critical noun phrase (e.g. 'the big pig'), as well as their reaction times for clicking the chosen object.

2. Theoretical background

Previous empirical research on perspective taking in narratives has investigated, among other things, the integration of information that is perceptually available to the character in the reader's mental model. For example, Horton and Rapp (2003) found that readers were slower to answer verification questions about objects that were visually inaccessible to story characters, compared to situations where such information was not occluded. This suggests that readers take into account characters' perceptual perspectives during narrative comprehension.

However, it is not yet clear where in the time course of narrative comprehension perspective taking takes place. Research on perspective taking in conversation suggests that listeners may not immediately take the speaker's perspective into account during language comprehension. For example, results from referential communication tasks have shown that listeners initially consider objects that they know cannot be seen by the speaker as possible referents of a speaker's referring expression (e.g. Keysar et al., 2000; Keysar et al., 2003). These results gave rise to the *perspective adjustment model* (e.g. Epley et al., 2004; Horton & Keysar, 1996), according to which core language comprehension processes are egocentric, while perspective taking is a separate cognitive process that only comes into play in an optional later adjustment stage of language processing. Other research has stressed that all kinds of contextual cues may influence language comprehension from the start of processing, including inferred mental states of other people (e.g. Brennan & Hanna, 2009). According to such *constraint-based models* of language processing (e.g. Brown-Schmidt & Hanna, 2011; Hanna et al., 2003), it depends on the strength of the different cues whether another person's perspective is immediately considered. Recently, it has been proposed that there may actually be two distinct cognitive processes taking into account speaker-specific information: an early automatic cue-driven process, and a later mentalizing process (Kronmüller & Guerra, 2020).

Taking the perspective of a character in narrative discourse appears to be a rather different phenomenon from taking the perspective of an interlocutor in a conversation. Still, both types of perspective taking may overlap in the cognitive processes involved, as both rely on Theory of Mind abilities (e.g. Apperly, 2010; Ferguson et al., 2015) and the suppression of a default (ego- or narrator-centric) perspective (Köder et al., 2015). Therefore, it is conceivable that perspective taking in narrative comprehension, too, involves a later adjustment process requiring mentalizing abilities (in line with the *perspective adjustment model*), and/or is characterized by cue-driven competition between the character's perspective and the comprehender's own perspective during processing (in line with the *constraint-based model*).

Narratives often contain contextual cues that may influence the extent to which readers take a character's perspective. Narratologists and theoretical linguists typically point to a few distinct grammatical devices of reported speech and thought as encoding such shifts from the narrator to the character perspective (e.g., Banfield, 1982; Fludernik, 2003; Schlenker, 2004; Eckardt 2014). Researchers in cognitive linguistics and stylistics, by contrast, favor a more fine-grained and inclusive approach, distinguishing a variety of more or less subtle cues for perspective shifting and different varieties and degrees of perspective mixing and blending (e.g., Vandelanotte, 2009; Dancygier, 2011; Dancygier et al., 2016). In this paper, we restrict attention to two specific grammatical report constructions for which it is unanimously agreed that they induce (a high degree of) perspective shifting: DD and FID. In DD (1b), the (internal) speech of the third-person protagonist ('she') is quoted. The FID variant in (1c) lacks the reporting frame. Nonetheless, it is still recognizable as a report of what a third-person story character is thinking in that, while the verb tense reflects the neutral point of view of the narrator, everything else reflects the subjective point of view of the character. The presence of expressive elements and constructions in the context of a third-person narrative is considered a signal for readers to shift to the character's perspective (Banfield, 1982; Fludernik, 2003). There is also some experimental evidence that confirms that expressive FID cues like epithets (e.g. 'the jerk') can boost shifting to the perspective of an individual other than the speaker/narrator (Kaiser, 2015).

The effect of different perspective-shifting devices on readers' narrative perspective taking has been investigated empirically. In a questionnaire study, Salem et al. (2017) explored the extent to which the narrative mode, e.g., first-person narration or FID, increased the likelihood of imagining the story from the character's spatial perspective. According to their findings, FID did not affect this likelihood. In a different study, Sato et al. (2012) compared omniscient third-person narration, which includes descriptions of characters' mental states, with "objective" third-person narration, in which only situations that can be externally observed are described. They found that narrative style affects whether readers imagine a scene from an external perspective or from "within" the character's perspective. However, there are few studies on the real-time processing of perspective shifts and specifically of FID (see Salem et al., 2018; Meuser et al., 2020). In this paper, we explore the real-time unfolding of narrative perspective shifting in DD and FID. In addition to investigating how readers ultimately resolve perspective-sensitive referring expressions, we crucially also use real-time measures to look into the cognitive process of reference resolution in perspective-shifting environments.

3. Predictions

We hypothesize that listeners do not automatically take the perspective of a narrative character when processing perspective-ambiguous referring expressions: On the one hand, different perspectives may be in competition, and listeners may eventually interpret the referring expression from their own or the narrator's perspective. On the other hand, perspective taking may be a late cognitive process, in which case the referring expression is initially processed from the listener's or narrator's perspective, even if the character's perspective is eventually selected. We therefore predict that, in the absence of explicit linguistic cues, listeners will often select the object that matches the description in the referring expression from their own or the narrator's perspective. If they do eventually select the object matching the character's perspective, we predict that the presence of a competing perspective will cause longer reaction times, and mouse movements will initially go towards the object matching the listener's or narrator's perspective. We further hypothesize that perspective-shifting cues that mark FID and DD will boost the listener's disposition to adopt the character's perspective: as compared to ND, listeners will be more likely to resolve perspective-ambiguous referring expressions in DD and FID to referents that are visually accessible to the character whose thoughts are being reported. In addition, since the reporting frame ('she thought') in DD makes perspective shifting explicit, we take it to be a stronger cue for perspective shifting than FID, which lacks this overt frame. Hence, we hypothesize that character-oriented interpretations of referring expressions are easier to get in DD than in FID. We therefore predict listeners to have a more direct mouse trajectory towards the character-accessible referent and shorter reaction times in the DD than in the FID condition, which in turn has straighter and faster trajectories compared to ND.

4. Materials and methods

4.1 Participants

Sixty-five participants (40 males, 24 females, 1 non-binary; mean age 29.1 years, range 18-75) were recruited via Prolific. Fifty participants used an external mouse, while 15 used a touchpad or touchscreen. Three participants were left-handed. Fourteen participants reported playing video games every day, 16 never played video games, with the remainder being occasional players. Participants were paid 4 GBP (~4.62 EUR) for their participation.

4.2 Materials

Twenty-four sets of visual stimuli were created, drawn by the second author. Each stimulus consisted of a narrative character in the center of the scene, in front of a barrier (a brick wall, a paneled wall or a paravan) obscuring the character's view behind them. Six objects were depicted around the character. In the Shared condition, three objects of the same type but of different sizes, i.e. a big one, a middle-sized one and a small one, were placed in front of the character, such that they were all in the character's view. In the Non-shared condition, one of these objects (either the biggest or the smallest one, depending on the expression used in the story) was placed in the middle slot behind the barrier. Left and right position of the character-accessible and character-inaccessible objects was counterbalanced across items. The remaining three slots were filled by distractor objects that were always of the same type and the same size.

Each visual stimulus was accompanied by an audio fragment of a narrative in the past tense, recorded by the last author. The fragment consisted of four sentences (example (1)). The first sentence introduced the character, avoiding to suggest any movement by the character, as that might have suggested that the character had just come from behind the barrier. The second sentence introduced the relevant objects by a plural definite noun phrase. The third sentence was the critical sentence, referring to one of the objects with a modified definite noun phrase (either 'the small X' or 'the big X'). In the FID condition, this sentence additionally started with an exclamative interjection (e.g. 'uh oh!'), and the sentence was read in a more emphatic voice, while retaining the past tense of the finite verb. The DD condition made this sentence a full quotation, starting with a reporting clause (e.g. 'she thought'), followed by a shift in prosody and the finite verb in the present tense. The final sentence in each narrative was a varying closure statement, intended to give the participant time to move their mouse until they found the correct object.

We also created 27 fillers in order to distract participants from the purpose of the experiment, three practice items, and four "calibration" items, which were intended to clarify to the participants that the objects behind the obstacle were hidden from the character's view.

4.3 Procedure

The experiment was programmed in PennController for Ibex (Zehr & Schwartz, 2018; https://doc.pcibex.net/), and conducted on the web. Participants were recruited via Prolific, where they were first asked to provide their informed consent. Next, they were asked what type of mouse they were using (mouse or trackpad).[1] Participants were then told that they would listen to short stories while looking at scenes depicting a human character and several objects. They were instructed to click on the object they thought the story was about as quickly as possible and to choose intuitively by moving their mouse immediately as soon as they knew which object was referred to, before the end of the story. Participants first went through the "calibration" and practice phases before the main experiment started.

Each trial started with the visual scene being presented. After a delay of 1500 ms, which was implemented to encourage participants to briefly inspect the scene, the start button appeared at the bottom of the screen. As soon as the participant clicked this button, the audio was played and the participant's mouse movements were tracked. Each clickable object was highlighted by a semi-transparent colored square. As soon as participants clicked an object, a dashed outline appeared around it. Participants could still change their mind and click on a different object until the end of the audio. They received a warning message if they clicked an object before the critical noun phrase had been uttered or after the audio had already finished. Fifteen seconds after the start button was clicked, the experiment progressed automatically to the next trial.

At the end of the experiment, participants were asked what they thought the experiment investigated. Only four participants wrote they thought it was about taking the perspective of the character. The whole experiment took about 20 minutes on average.

4.4 Design and analysis

Crossing the factors Sharedness (Shared, Non-shared) and Discourse type (ND, FID, DD) resulted in a 2x3 within-participants and within-items design. The items were distributed over six lists according

[1] Use of a phone or tablet was blocked by the software.

to a Latin Square design, such that each item on a list was presented in only one condition. Within a list, items were ordered quasi-randomly. Participants were randomly assigned to one of the lists.

We investigated the effects of our conditions on several measures: First, we examined which object (character-accessible or character-inaccessible) was clicked.[2] Second, we measured the reaction time (RT) from the offset of the critical NP to the selection of the chosen object. Third, we investigated three common mouse tracking measures (Freeman & Ambady, 2010): (a) the Area Under the Curve (AUC), which is the geometric area between the actual trajectory and a straight line from start to end point; (b) the maximum absolute deviation (MAD), which is the furthest perpendicular distance from the actual trajectory to a straight line from start to end point; and (c) the number of directional changes along the x-axis (x-flips). Both AUC and MAD are measures of spatial attraction towards an object, and hence are informative of the amount of competition between potential referents (Freeman & Ambady, 2010; Tomlinson et al., 2013). The number of x-flips is a measure of complexity within the trajectory, and may also be informative of competition (Freeman & Ambady, 2010).

We analysed our data in R 4.0.3 (R Core Team, 2020), using the *mousetrap* package (Kieslich & Henninger, 2017) and the *lme4* package (Bates, Mächler et al. 2015). Object selection was analyzed using logistic mixed effects models. For the RT and mouse tracking analyses, selections of the character-accessible object and character-inaccessible object were analyzed separately. Log-transformed RTs, AUC and MAD were analyzed using linear mixed effects models, and x-flips were analyzed using poisson mixed effects models. Sharedness and Discourse type were always included as fixed factors, and were centered to reduce collinearity. We started out with by-participant and by-item random intercepts as well as by-participant and by-item random slopes for Sharedness and Discourse type, aiming at the maximal random effects structure justified by the data (Barr et al., 2013). In case of non-convergence, we used model selection to simplify the random effects structure, starting with random slopes that had the lowest variance (Bates, Kliegl et al., 2015). The best model was determined on the basis of model comparison using a likelihood ratio test and AIC. Significance of predictors in the linear mixed effects models was determined using the lmerTest package in R (Kuznetsova et al., 2017).

5. Results

We excluded trials in which an object was selected before the offset of the critical NP (N=193). We additionally excluded trials where object selection occurred either above or below 2.5 standard deviations from the mean log-transformed reaction time of a character-accessible or a character-inaccessible object selection (N=30). Finally, we excluded trials in which participants selected a distractor object (N=4) or in which no object was selected before the end of the audio (N=2). This left 1331 object selections for analysis.

5.1 Object selection

Figure 2 shows the proportion of selections of the character-accessible and character-inaccessible objects. If participants would always shift to the character's perspective in the narratives, one would expect them to always click on a character-accessible object. Clearly, this was not the case: in the Non-shared condition, participants selected the character-inaccessible object about 60% of the time. Surprisingly, neither an FID nor a DD context seemed to influence this choice. This was confirmed by a non-significant effect of Discourse type (DD: p =.49 , FID: p =.33) on the likelihood of selecting the character-accessible object.[3]

---

[2] We only counted first selections, i.e. if participants clicked more than once within a trial, we ignored the later selections. The distribution of final selections was very similar to that of first selections.

[3] We only included choices from the Non-Shared condition (N=648), given that in the Shared condition there was no character-inaccessible object. Our final model included only random intercepts for participants and items.

Figure 2. Proportion of clicks on the character-accessible and character-inaccessible object, across the Shared and Non-shared conditions and the three discourse contexts.

5.2 Reaction times

Figure 3 shows the mean RTs from the offset of the critical NP, for selections of the character-accessible object (N=936) in the Shared and the Non-shared condition, and the mean RTs for selections of the character-inaccessible object in the Non-shared condition (N=395). The final model included the fixed factors Sharedness and Discourse type and their interaction, by-participant and by-item random intercepts, and a by-participant random slope for Sharedness. Although RTs in the Non-shared condition were numerically longer than in the Shared condition, in line with our predictions, this difference did not reach significance (p = .08). However, there was a significant main effect of Discourse type: DD had significantly shorter reaction times than ND (β = -0.18, SE = 0.09, t = -2.09 , p = .04), while FID did not seem to influence participants' reaction times significantly (p = .21). This suggests that the overt perspective-shifting cues of DD facilitated listeners to shift to the character-accessible object. The interaction between Sharedness and Discourse type was not significant (FID vs ND: p = .48; DD vs ND: p = .85), suggesting that the effect of the perspective-shifting cues did not depend on whether perspective between listener/narrator and character was shared or not.

In the character-inaccessible object selections, Discourse type did not have a significant effect on RT (DD: p = .44, FID: p = .89), indicating that DD and FID did not affect ease of processing (rightmost graph of Figure 3).

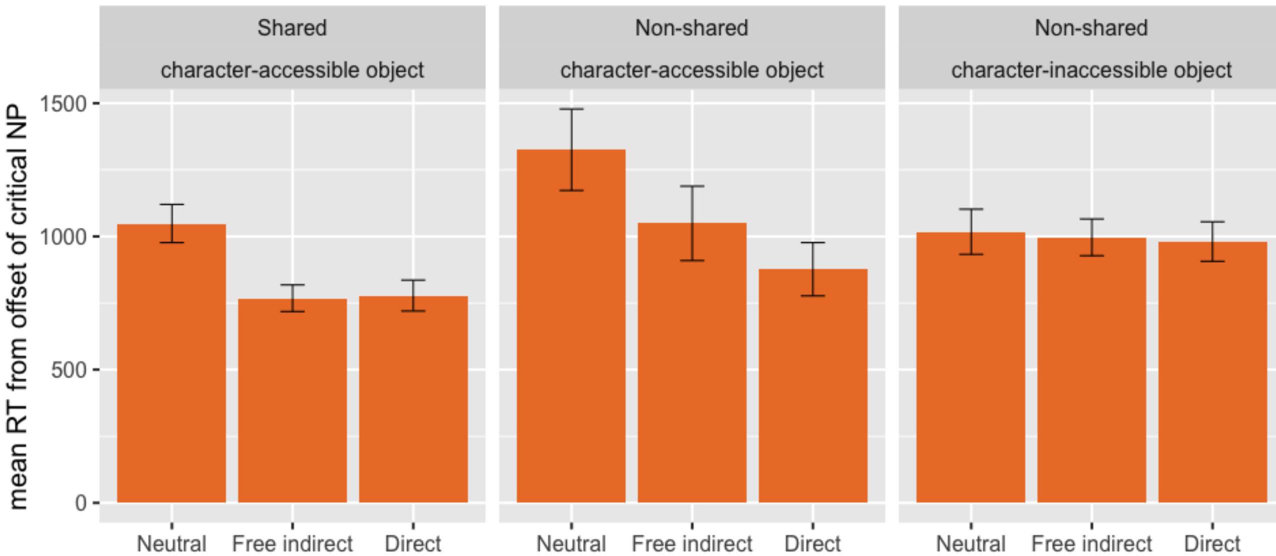

Figure 3. Mean RTs from the offset of the critical NP to the selection of the character-accessible or character-inaccessible object, by Sharedness and Discourse type conditions. Error bars reflect standard errors.

5.3 Mouse trajectories

Mouse trajectories were extracted from the raw data, keeping only mouse positions between the offset of the critical NP and object selection. Since participants' screen resolutions varied, we scaled all mouse coordinates to a 1056x477 pixel canvas. Next, all trials with fewer than two logged mouse positions (N=135) were excluded, leaving 801 trials for analysis. Trajectories towards the character-accessible object were remapped to the left side, and trajectories towards the character-inaccessible object were remapped to the right side. Finally, the trajectories were time-normalized (Kieslich et al., 2018; Spivey et al., 2005), such that there was an equal number of time stamps (101) for each trajectory. The resulting mouse trajectories for the trials in which participants eventually selected the accessible object, averaged by participant, are shown in Figure 4. The patterns for the different conditions clearly suggest more uncertainty in the Non-shared condition (top of Figure 4) than in the Shared condition (bottom of Figure 4). Furthermore, at the offset of the critical NP in the Non-shared condition, the mouse is closer to the eventually selected object in DD than in FID, and much closer than in ND. This suggests that the initial DD cues help participants to already move their mouse towards the character-accessible part of the scene. These participants seemed to even be sensitive to the more mixed cues of FID. The mouse-tracking starting points in the Non-shared condition in Figure 4 thus show the same pattern as the RTs in Figure 3. After the starting points we seem to also see most uncertainty in ND, then FID, then DD, suggesting that the perspective-shifting cues indeed help participants quickly select the character-accessible object.

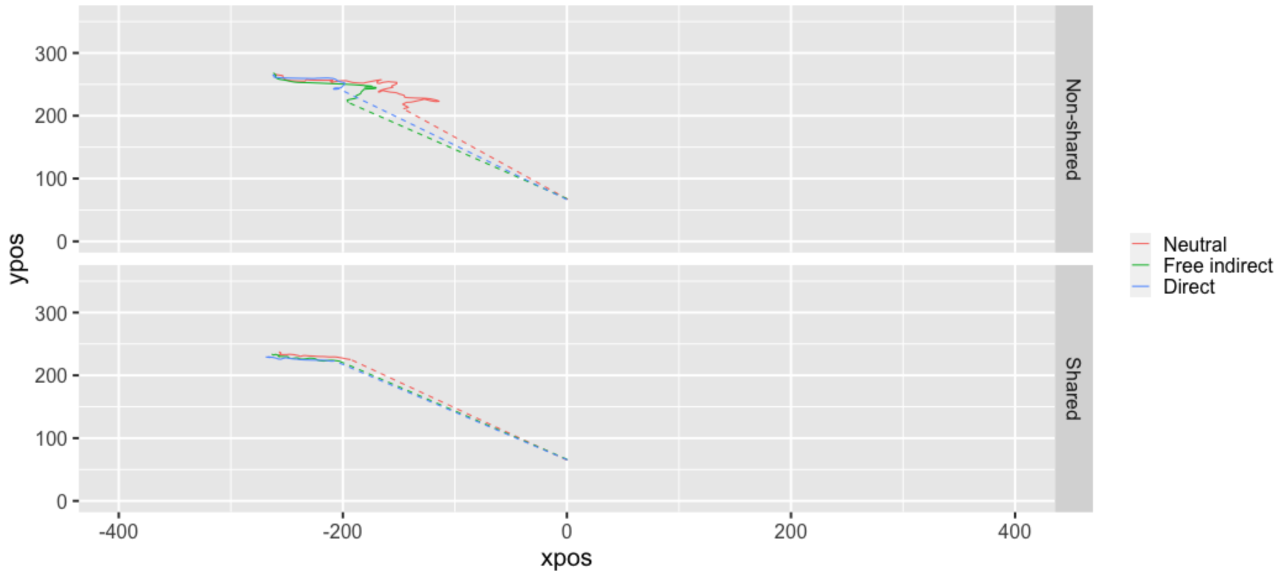

Figure 4. Time-normalized mouse trajectories in trials where participants selected the character-accessible object, from the offset of the critical NP to the first object selection, by Sharedness and Discourse type condition. Dashed lines represent the excluded mouse movements from the start of mouse tracking until the offset of the NP.

The analyses of the three mouse-tracking measures showed a mixed pattern, however. AUC and MAD did not show significant effects of either Sharedness or Discourse type, probably because variance was very large. The number of x-flips, on the other hand, did reveal a significant effect of Sharedness ($\beta = 0.49$, $SE = 0.20$, $t = 2.50$, $p = .01$): there were more directional changes along the x-axis in the Non-shared (1.1 on average) than in the Shared condition (0.6 on average). This main effect was qualified by a significant interaction with Discourse type ($\beta = -0.56$, $SE = 0.27$, $t = -2.08$, $p = .04$), suggesting that the effect of Sharedness was driven by the ND condition, which showed 1.5 x-flips in the Non-shared condition, as compared to 0.9 in both FID and DD conditions.

Because the overall number of x-flips was still rather low in this short time frame, we also examined a more fine-grained measure of uncertainty in the mouse trajectories: sample entropy. Sample entropy quantifies the certainty with which sequences of mouse positions along the x-axis can be predicted given previous sequences, with higher values expressing more uncertainty and 0 expressing absolute certainty (Hehman et al., 2015). The results for sample entropy looked very similar to those of the number of x-flips, with higher entropy in the Non-shared condition (0.06) than in the Shared (0.04), especially in the ND condition (Non-shared: 0.08). The main effect of Sharedness was significant ($\beta = 0.03$, $SE = 0.01$, $t = 2.48$, $p = .02$), but the interaction with Discourse type did not reach significance (FID vs ND: $p = .12$; DD vs ND: $p = .46$). Together, these results suggest that participants were more uncertain when the best-fitting referent was visually inaccessible to the character than when all possible referents were shared between listener and character, and possibly more so when no overt cues were present to suggest that the story should be interpreted from the character's perspective.

6. Discussion

We conducted a web-based mouse-tracking study in order to explore the time course of resolving perspective-ambiguous referring expressions in narratives, and investigated how different perspective-shifting cues affect listeners' narrative perspective taking. Participants' offline responses confirmed our hypothesis that referring expressions in narratives are often interpreted from the narrator's or the listener's own visual perspective: participants selected the character-inaccessible object 60% of the time in the Non-shared condition. In those cases in which participants did select the character-accessible object, reaction times were numerically (but not significantly) longer in the Non-shared than in the Shared condition, and the Non-shared condition showed more uncertainty in the mouse trajectories than the Shared condition. These results indicate that the observer's perspective competes with the character's perspective during reference processing.

We further hypothesized that the perspective-shifting cues included in our FID and DD conditions would facilitate a character-oriented interpretation of the referring expressions. Surprisingly, however, the presence of FID or DD cues did not increase the proportion of character-accessible object selections. Still, the RT results showed that DD cues caused participants selecting the character-accessible object to make their selection significantly faster, indicating that perspective-shifting cues can facilitate reference processing by guiding the listener to a set of character-accessible referents. The FID condition did not have the same effect, which is in line with FID being a mixed form of direct and indirect discourse features.

Our mouse-tracking analyses of character-accessible object selections showed a slightly different pattern: the uncertainty in the Non-Shared condition (at least as evidenced by the increase in x-flips) was more pronounced in the ND condition, where perspective-shifting cues were absent, than in both the FID and the DD conditions, suggesting that people were also aided by the more mixed perspective-shifting cues. Even so, there seemed to be quite some individual variation, especially in our analyses of AUC and MAD.[4] To investigate this individual variation further, we briefly examined whether we could detect separate strategies in individual participants. We observed that nine out of 65 participants never selected the inaccessible object in the Non-shared condition, whereas 20 participants consistently chose the inaccessible object. This suggests that almost half of the participants employed a particular strategy for handling the task: either consistently take the character's perspective or pick the objectively best-fitting referent. It is possible that the participants that exclusively chose the character-accessible objects may not be sensitive to the different perspective-shifting cues, precisely because they tend to always take the character's perspective. The rather large group of participants who always selected the objectively best-fitting referent might not have paid attention to the obstacle or considered all depicted objects as visible to the character.[5] Alternatively, they might have seen the referential task as a game, in which they had to find the "correct answer", and in which any perspective-taking cues were considered distractors to mislead them in finding the objectively correct answer. If this is the case, it is interesting to look at the mouse trajectories for the inaccessible object selections, to see if participants who ignored the character's perspective may still be sensitive to the presence of perspective-shifting cues. Here, the results for AUC and MAD again do not show any significant effects, and neither does the number of x-flips. Sample entropy does show that mouse movements towards the inaccessible object are significantly more uncertain in the FID than in the ND condition. However, the same does not hold for DD, so it remains unclear how to interpret these results. We are currently investigating individual strategies in perspective taking more systematically, using a larger group of participants.

One potential confound in our stimuli is that because of the additional perspective-shifting cues, the auditory stimuli were longer in the FID and DD conditions than in the ND condition. Hence,

---

[4] One possibility is that this large variance partly has to do with the use of different mouse devices. Fifteen out of 65 participants used a laptop touchpad or a touchscreen which may have affected the mouse trajectories. However, excluding those fifteen participants resulted in quite similar average mouse trajectories.

[5] We used participants' responses to the 'calibration' items and the fillers that were completely unambiguous to check whether some participants might not have done the task seriously (e.g. by always clicking the same selection box without attending to the stimuli). All participants had high accuracy scores on these items, so it is unlikely that such behavior contaminated our results.

participants had more time to find the correct referent in these conditions, which could potentially explain their shorter RTs after the offset of the critical NP. However, such an effect should occur across the board, irrespective of which object was selected. As Figure 3 shows, no facilitation effect of the FID and DD conditions was found when participants selected the character-inaccessible object, suggesting that the effect in the character-accessible object selections was related to our perspective manipulation.

In all, our findings provide some support for the hypothesis that listeners can integrate cues about the perspective of narrative characters quite quickly in their processing of referring expressions: even before any overt linguistic cues were presented, participants' mouse movements were already leaning towards the side of the scene accessible to the character, and the presence of especially direct discourse cues only reinforced this tendency. At the same time, mouse movements still showed some increased uncertainty in an ambiguous referential context without any overt perspectival cues, which suggests that the character's perspective received competition from the observer's perspective. These findings are more in line with models of language processing that allow for the influence of all kinds of contextual and pragmatic factors in the earliest stages of language processing (e.g. Brennan & Hanna, 2009), than with models that assume that perspective information only comes into play in a late processing stage (e.g. Epley et al., 2004; Kronmüller & Guerra, 2020). Still, it is striking that there were many cases in which referring expressions were resolved to the character-inaccessible object, suggesting that perspective information was ignored altogether. Future research should reveal whether this is a task effect, or whether this is a genuine language processing strategy employed by certain comprehenders.

Furthermore, our results are compatible with research stressing the facilitating role of DD in narrative contexts (Köder & Maier, 2018). Findings for FID are less conclusive regarding its role in perspective shifting. This may be due to large individual variation regarding the sensibility to literary cues such as FID, and is in line with views regarding this kind of construction as ambiguous (see Bray, 2007). Yet, we believe that evidence from the mouse trajectory patterns reveals a processing difference that can be rendered more transparent in future research with further methodological refinements.

Funding information

This research was supported by the Netherlands Organisation for Scientific Research (NWO), under Grant 275-89-0360 to the first author, and 276-80-004 (FICTION) to the second and last authors.

Acknowledgments

We thank the members of the Semantics and Cognition group at the University of Groningen, the attendees of the Dutch Annual Language Day 2021, and two anonymous reviewers for their helpful comments on earlier versions of this work.

References

Apperly, Ian. 2010. *Mindreaders: The Cognitive Basis of "Theory of Mind."* London: Psychology Press. https://doi.org/10.4324/9780203833926

Banfield, Ann. 1982. *Unspeakable sentences: Narration and representation in the language of fiction*. Boston: Routledge & Kegan Paul Ltd.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. "Random effects structure for confirmatory hypothesis testing: Keep it maximal." *Journal of memory and language* 68 (3): 255–278. https://doi.org/10.1016/j.jml.2012.11.001)

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. "Parsimonious mixed models." arXiv preprint arXiv:1506.04967.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01

Bray, Joe. 2007. “The ‘dual voice’ of free indirect discourse: a reading experiment.” *Language and Literature*. Sage Publications Sage CA: Thousand Oaks, CA 16 (1): 37–52. https://doi.org/10.1177/0963947007072844

Brennan, Susan E. & Joy E. Hanna. 2009. “Partner-specific adaptation in dialog.” *Topics in Cognitive Science* 1 (2): 274–291. https://doi.org/10.1111/j.1756-8765.2009.01019.x

Brown-Schmidt, Sarah & Joy Hanna. 2011. “Talking in another person’s shoes: Incremental perspective-taking in language processing.” *Dialogue & Discourse* 2 (1): 11–33. https://doi.org/10.5087/dad.2011.102

Dancygier, Barbara. 2011. *The Language of Stories: A Cognitive Approach*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511794414

Dancygier, Barbara, Wei-lun Lu & Arie Verhagen, eds. 2016. *Viewpoint and the Fabric of Meaning* (Form and Use of Viewpoint Tools across Languages and Modalities, 10). Berlin: Mouton.

Eckardt, Regine. 2014. *The semantics of free indirect speech. How texts let you read minds and eavesdrop*. Vol. 31. Leiden: Brill.

Epley, Nicholas, Boaz Keysar, Leaf Van Boven & Thomas Gilovich. 2004. “Perspective Taking as Egocentric Anchoring and Adjustment.” *Journal of Personality and Social Psychology* 87 (3): 327–339. https://doi.org/10.1037/0022-3514.87.3.327

Ferguson, Heather J., Ian Apperly & Jumana Ahmad, Markus Bindemann & James Cane. 2015. “Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task.” *Cognition* 139: 50–70. https://doi.org/10.1016/j.cognition.2015.02.010

Fludernik, Monika. 2003. *The fictions of language and the languages of fiction*. London, New York: Routledge.

Freeman, Jonathan B. & Nalini Ambady. 2010. “MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method.” *Behavior Research Methods* 42 (1): 226–241. https://doi.org/10.3758/BRM.42.1.226

Hanna, Joy E, Michael K Tanenhaus & John C Trueswell. 2003. “The effects of common ground and perspective on domains of referential interpretation.” *Journal of Memory and Language* 49 (1): 43–61. https://doi.org/10.1016/S0749-596X(03)00022-6

Hehman, Eric, Ryan M. Stolier & Jonathan B. Freeman. 2015. “Advanced mouse-tracking analytic techniques for enhancing psychological science.” *Group Processes & Intergroup Relations* 18 (3): 384–401. https://doi.org/10.1177/1368430214538325

Heller, Daphna, Daniel Grodner & Michael K. Tanenhaus. 2008. “The role of perspective in identifying domains of reference.” *Cognition* 108 (3): 831–836. https://doi.org/10.1016/j.cognition.2008.04.008

Horton, William S. & Boaz Keysar. 1996. “When do speakers take into account common ground?” *Cognition* 59 (1): 91–117. https://doi.org/10.1016/0010-0277(96)81418-1

Horton, William S. & David N. Rapp. 2003. “Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension.” *Psychonomic Bulletin & Review* 10 (1): 104–110. https://doi.org/10.3758/BF03196473

Kaiser, Elsi. 2015. “Perspective-shifting and free indirect discourse: Experimental investigations.” *Proceedings of Semantics and Linguistic Theory* 25: 346–372. https://doi.org/10.3765/salt.v25i0.3436

Keysar, Boaz, Dale J. Barr, Jennifer A. Balin & Jason S. Brauner. 2000. “Taking perspective in conversation: The role of mutual knowledge in comprehension.” *Psychological Science* 11 (1): 32–38. https://doi.org/10.1111/1467-9280.00211

Keysar, Boaz, Shuhong Lin & Dale J Barr. 2003. "Limits on theory of mind use in adults." *Cognition* 89 (1): 25–41. https://doi.org/10.1016/S0010-0277(03)00064-7

Kieslich, Pascal J. & Felix Henninger. 2017. "Mousetrap: An integrated, open-source mouse-tracking package." *Behavior Research Methods* 49 (5): 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

Kieslich, Pascal J., Felix Henninger, Dirk U. Wulff, Jonas Haslbeck & Michael Schulte-Mecklenbeck. 2018. "Mouse-tracking: A practical guide to implementation and analysis." PsyArXiv. https://doi.org/10.31234/osf.io/zuvqa

Köder, Franziska & Emar Maier. 2018. "The advantage of story-telling: children's interpretation of reported speech in narratives." *Journal of Child Language*. Cambridge University Press 45 (2): 541–557. https://doi.org/10.1017/S0305000917000344

Köder, Franziska, Emar Maier & Petra Hendriks. 2015. "Perspective shift increases processing effort of pronouns: a comparison between direct and indirect speech." *Language, Cognition and Neuroscience*. Routledge 30 (8): 940–946. https://doi.org/10.1080/23273798.2015.1047460

Kronmüller, Edmundo & Ernesto Guerra. 2020. "Processing speaker-specific information in two stages during the interpretation of referential precedents." *Frontiers in Psychology* 11. https://doi.org/10.3389/fpsyg.2020.552368

Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models" *Journal of Statistical Software* 82: 1-26.

Meuser, Sara, Stefan Hinterwimmer & Maximilian Hörl. 2020. "Online-processing of protagonists' perspective-taking" (poster presented at CUNY 2020).

Nadig, Aparna S. & Julie C. Sedivy. 2002. "Evidence of Perspective-Taking Constraints in Children's On-Line Reference Resolution." *Psychological Science* 13 (4): 329–336. https://doi.org/10.1111/j.0956-7976.2002.00460.x

R Core Team. 2020. "R: A language and environment for statistical computing." *R foundation for statistical computing*. Vienna, Austria, https://www.R-project.org/

Salem, Susanna, Thomas Weskott & Anke Holler. 2017. "Does narrative perspective influence readers' perspective-taking? An empirical study on free indirect discourse, psycho-narration and first-person narration." *Glossa: a journal of general linguistics* 2 (1): 61. https://doi.org/http://doi.org/10.5334/gjgl.225

Sato, Manami, Hiromu Sakai, Jennifer Wu & Benjamin K. Bergen. "Towards a Cognitive Science of Literary Style: Perspective-Taking in Processing Omniscient versus Objective Voice." *Proceedings of the 34th Annual Conference of the Cognitive Science Society* ed. by Naomi Miyake, David Peebles and Richard P. Cooper, 959–964. Austin, TX: Cognitive Science Society.

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

Spivey, Michael J., Marc Grosjean & Günther Knoblich. 2005. "Continuous attraction toward phonological competitors." *Proceedings of the National Academy of Sciences*. National Academy of Sciences 102 (29): 10393–10398. https://doi.org/10.1073/pnas.0503903102

Tomlinson, John M., Todd M. Bailey & Lewis Bott. 2013. "Possibly all of that and then some: Scalar implicatures are understood in two steps." *Journal of Memory and Language* 69 (1): 18–35. https://doi.org/10.1016/j.jml.2013.02.003

Vandelanotte, Lieven. 2009. *Speech and Thought Representation in English: A Cognitive-Functional Approach*. https://doi.org/10.1515/9783110215373

Zehr, Jeremy & Florian Schwartz. 2018. "PennController for Internet Based Experiments (IBEX)." https://doi.org/10.17605/OSF.IO/MD832

Supplementary data

The materials and raw data from this study can be found here:
https://doi.org/10.17605/OSF.IO/3R7BQ