



# The Roles of Implicit Causality and Discourse Context in Pronoun Resolution

Emiel van den Hoven\* and Evelyn C. Ferstl

Center for Cognitive Science, University of Freiburg, Freiburg, Germany

Some interpersonal verbs show a bias in the proportion of times their subject and object arguments are rementioned in a sample of explanations for the eventuality the verb describes. This bias is known as the implicit causality bias. Several studies have shown that readers and listeners rapidly use the implicit causality bias during pronoun resolution. Whether listeners also rapidly incorporate relevant contextual information during pronoun resolution, is an open question. In the current paper, we report two visual world eye-tracking studies intended to answer this question. Participants listened to stories that included implicit causality verbs followed by a “because” clause with an ambiguous pronoun in its subject position. During the story, the participants looked at a screen on which potential referents of the ambiguous pronoun were displayed. In Experiment 1, a simple main effect of implicit causality bias on looks toward the character that was congruent with the bias was found among items in one of the two discourse conditions. Discourse context, however, only affected looks for a subset of verbs and in the opposite direction of what was hypothesized. In Experiment 2, no main effects of IC Bias or discourse context were found, but there was a marginally significant interaction which was not hypothesized. In both experiments, discourse context influenced looks only for a subset of verbs and never in the predicted direction. The results favor an account in which the influence of lexical semantics is, at least initially, stronger than the influence of world knowledge, and discourse context. Additional exploratory analyses suggested that eye movements already reveal remention biases at an early point in the sentence, whereas the causal potency of the subject argument is predicted by looks starting from the onset of the causal connective.

**Keywords:** implicit causality, discourse context, pronoun resolution, visual world paradigm, eye-tracking

## OPEN ACCESS

### Edited by:

Ping Li,  
Pennsylvania State University,  
United States

### Reviewed by:

Joshua Hartshorne,  
Harvard University, United States  
Alan Garnham,  
University of Sussex, United Kingdom

### \*Correspondence:

Emiel van den Hoven  
emiel.vandenhoven@  
frequenz.uni-freiburg.de

### Specialty section:

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

**Received:** 19 April 2018

**Accepted:** 13 November 2018

**Published:** 04 December 2018

### Citation:

van den Hoven E and Ferstl EC (2018)  
The Roles of Implicit Causality and  
Discourse Context in Pronoun  
Resolution. *Front. Commun.* 3:53.  
doi: 10.3389/fcomm.2018.00053

## THE ROLES OF DISCOURSE CONTEXT AND IMPLICIT CAUSALITY IN PRONOUN RESOLUTION

Pronoun resolution is not only guided by morphosyntactic constraints, such as gender, number, and person, but also by soft constraints, such as the first-mention preference (Frederiksen, 1981; Gernsbacher and Hargreaves, 1988), subject preference (Crawley et al., 1990); grammatical parallelism (Sheldon, 1974; Smyth, 1994) and, most important for our current purposes, the implicit causality (IC) bias (Garvey and Caramazza, 1974).

The IC bias is the bias, shown by interpersonal verbs, in the proportion of times their subject and object arguments are rementioned in a sample of explanations for the eventuality that the verb describes. When a group of participants is asked to complete sentences like

(1a) and (1b), the verb *amaze* is associated with a preference for explanations about the first noun phrase (so *amaze* is an NP<sub>1</sub> biased verb), whereas the verb *love* is associated with a preference for explanations about the second noun phrase (so *love* is an NP<sub>2</sub> biased verb; Ferstl et al., 2011).

(1a) Sally amazed Mary because she...

(1b) Sally loved Mary because she...

IC is a soft constraint in the sense that sentence continuations that are incongruent with the bias are not ungrammatical, but merely less common and generally more difficult to devise and interpret. However, some researchers (e.g., Crinean and Garnham, 2006; Guerry et al., 2006; Hartshorne et al., 2015) consider IC bias to be a function of hardcoded language structural elements rather than the result of inference processes involving world knowledge. On this account, the *lexical semantic account*, comprehenders rely on the verb's semantic structure in combination with a causal discourse relation to interpret an ambiguous pronoun, or to choose a topic with which to continue a sentence. World knowledge can ultimately influence pronoun resolution on the lexical semantic account, but only through revision of the initial interpretation (Hartshorne, 2014).

A competing account (e.g., Corrigan and Stevenson, 1994; Pickering and Majid, 2007; Van den Hoven and Ferstl, 2017) puts inference processes involving world knowledge center stage, in the tradition of Hobbs (1979). According to this *world knowledge account*, the verb's semantic structure is only a reliable predictor of IC bias insofar as verbs that share the same semantic structure (i.e., verbs from the same verb class) tend to evoke similar explanations. Semantic structure and IC bias are both products of conceptual knowledge.

The discourse context in which the sentences containing IC verbs are embedded has been shown to affect remention patterns (Van den Hoven and Ferstl, 2018). When a verb is used in an isolated sentence, comprehenders may make additional assumptions about the (fictional) situational context. For instance, when participants read the preamble "Marcel criticized Aaron because..." they are likely to imagine a situation in which the criticism is sincere. If this is the case, then a straightforward way to continue the sentence is by devising a reason for the criticism: What is it that Aaron did wrong? This leads to a preference in favor of rementioning the patient of *criticize*, making it an NP<sub>2</sub> verb.

However, in particular discourse contexts, the assumption of sincerity can be violated. Consider Version 1 of the story in **Table 1**. In this story, Marcel is seemingly in a legitimate position to criticize Aaron, since Marcel is portrayed as a successful artist, whereas Aaron is untalented. A straightforward way to continue this story, then, is to elaborate on what Aaron did that caused Marcel to criticize him (e.g., "Aaron gave the art school a bad reputation"). However, in Version 2, Marcel is portrayed as being less successful than Aaron. Here, based on the information given by the discourse, it may be inferred that Marcel is not sincere in his criticism, but merely criticizes Aaron out of spite. An explanation that conveys this information (e.g., "Marcel was jealous of Aaron") is likely to be prioritized over an explanation that conveys Marcel's purported reason for criticizing Aaron,

because it is more important to know that Marcel is not sincere than to know what the reason was that Marcel might have given for his criticism (see Van den Hoven and Ferstl, 2018). In other words, the Question Under Discussion (Roberts, 2012) has shifted from a question about an external reason to a question about an internal reason (Bott and Solstad, 2014). A shift in the priority of these kinds of information often leads to a shift in the remention bias, such that external reasons like those preferentially evoked by Version 1 lead to more NP<sub>2</sub> rementions than internal reasons like the ones preferentially evoked by Version 2.

Previous studies on implicit causality, employing different kinds of manipulations, showed only weak effects of world knowledge on rementions at best. Hartshorne (2014), for instance, found that the social status of the participants did not affect the interpretation of the pronoun: For a given verb, sentences like *The king criticized the knight because he...*, where the NP<sub>1</sub> has a higher social status than the NP<sub>2</sub>, were associated with the same remention bias as sentences like *The knight criticized the king because he...*, where the NP<sub>2</sub> has a higher social status. When participants were explicitly asked to judge whether the eventuality was due to the kind of person that the event participants were (e.g., how likely it was that the event took place because the king is the type of person who criticizes people), higher status event participants were rated as more causal than lower status event participants (see also Lafrance et al., 1997; Corrigan, 2001). And when an altered version of the remention task was used, such that participants indicated the referent of the pronoun in sentences like *The king criticized the knight because he is the kind of person that...*, there was a small effect of social status for a subset of verbs.

The gender of the event participants has been found to have a similarly small effect on rementions. Ferstl et al. (2011) found that men (but not women) were more likely to remention male event participants than female event participants in an explanation, and that male event participants were more likely than female event participants to be rementioned in an explanation for eventualities with negative emotional valence (e.g., *hit, kill, torment*). Hartshorne (2014) did not find an effect of the event participants' gender, except when a subset of verbs was used that showed a gender effect in Ferstl et al. (2011), Lafrance et al. (1997), using a more explicit causal attribution task, also found a preference for male event participants to be the initiator of events, particularly for events with negative valence.

A final factor that does not reliably affect rementions, although it does affect causality attribution, are the attitudes and behaviors of a set of alternative people. Majid et al. (2006) found that, when devising a continuation for a sentence like *Ellen pleased Paul because...*, participants do not take into consideration whether the discourse context states that there are few or many people who please Paul. What is important, however, is whether a negative quantifier (e.g., *few, not quite all*) or a positive quantifier (e.g., *a few, nearly all*) is used in the discourse context: Negative quantifiers led to more NP<sub>1</sub> rementions than positive quantifiers. Again, when the causality attribution task was made more explicit, and participants were asked whether there was

TABLE 1 | Two versions of an incomplete story.

| NP <sub>2</sub> biased story  | NP <sub>1</sub> biased story   |
|---|--|
| <p>Marcel fühlte sich <b>großartig</b>. Seit er vor drei Jahren seinen Abschluss an der Kunsthochschule gemacht hatte, hatte er <b>keine</b> Schwierigkeiten gehabt, Arbeit zu finden. Eines Tages ging er zum Arbeitsplatz eines alten Mitschülers, Aaron, um seine Bilder zu sehen. Wie sich herausstellte, war er recht <b>untalentierte</b>. Im Atelier kritisierte Marcel Aaron, weil...</p> | <p>Marcel fühlte sich <b>unangemessen</b>. Seit er vor drei Jahren seinen Abschluss an der Kunsthochschule gemacht hatte, hatte er <b>viele</b> Schwierigkeiten gehabt, Arbeit zu finden. Eines Tages ging er zum Arbeitsplatz eines alten Mitschülers, Aaron, um seine Bilder zu sehen. Wie sich herausstellte, war er recht <b>erfolgreich</b>. Im Atelier kritisierte Marcel Aaron, weil...</p> |
| <p><i>Marcel felt <b>great</b>. Since he graduated from art school three years ago, he had had <b>no</b> problems finding work. One day he went to the workplace of an old classmate, Aaron, to see his paintings. He turned out to be quite <b>untalented</b>. In the studio, Marcel criticized Aaron because...</i></p>   | <p><i>Marcel felt <b>inadequate</b>. Since he graduated from art school three years ago, he had had <b>a lot of</b> problems finding work. One day he went to the workplace of an old classmate, Aaron, to see his paintings. He turned out to be quite <b>successful</b>. In the studio, Marcel criticized Aaron because...</i></p>   |

*The story on the left is designed to create the expectation of an NP<sub>2</sub> remention, whereas the story on the right is designed to create the expectation of an NP<sub>1</sub> remention.*

something special about the NP<sub>1</sub> and NP<sub>2</sub>, there was a clear effect of set size.

In contrast to these findings from previous research showing a negligible influence of world knowledge on rementions, Van den Hoven and Ferstl (2018) did find a clear effect of the discourse context. A straightforward explanation for this discrepancy is the fact that in Van den Hoven and Ferstl's (2018) study, participants spent more time attending to the experimental manipulations than in the other studies we discussed above. In studies in which the event participants are manipulated, as well as in Majid et al.'s (2006) study, in which a single discourse context sentence was manipulated, the amount of lexical material that is manipulated is relatively small. In contrast, in Van den Hoven and Ferstl (2018), the discourse context consisted of more than three sentences on average, with on average more than four words differing between conditions. This simple fact may have made the information conveyed by lexical material other than the IC verb itself more salient.

A compatible explanation is that the types of manipulations used in Van den Hoven and Ferstl (2018) are more effective in biasing rementions than the manipulations used in previous research. Van den Hoven and Ferstl employed various types of discourse manipulations (see Materials), and most types had not been used in studies on IC before. Perhaps knowing whether the agent is sincere and/or well-informed is more relevant to remention biases than knowing the social status and gender of the event participants. Stories that involved the manipulation of covariation information (see Materials) allowed participants to develop a more detailed understanding of the story (or "situation model") than the single context sentence, itself devoid of context, that manipulated covariation information in Majid et al. (2006). Having a detailed understanding of the story situation, and a sense of how the covariation information naturally coheres with the eventuality in the main clause of the target sentence, may be essential for covariation information to affect rementions. More research is needed to disentangle the effect of the amount of lexical material used in the manipulations (which may simply be an effect of time-on-task) and the content of the manipulations. What is clear, however, is that IC remention biases can be modulated by the larger discourse context.

The finding that the larger discourse context influences remention biases in the context of IC verbs rules out any account of IC that denies the role of world knowledge altogether. However, on Hartshorne's (2014) lexical semantic account,

world knowledge does affect pronoun resolution, but only after an initial phase during which only language-structural factors are used to resolve the pronoun. After this initial phase, the interpretation of the pronoun can be revised with the help of world knowledge. The possibility of revision is necessary on any account that claims IC has an early effect on pronoun resolution, in order to allow for the comprehension of explanations that are incongruent with the verb's bias, as in "Sally<sub>i</sub> amazed Mary<sub>j</sub> because she<sub>j</sub> was easily impressed." Since the question of whether there is an initial processing stage during which language-structural information is privileged requires information about the time course of processing, it cannot be addressed using story completion studies. We here report two visual world eye-tracking studies intended to test the hypothesis that language-structural elements are privileged (henceforth the lexico-semantic hypothesis). First, however, we will review previous studies that have investigated the time course of the use of the IC bias during sentence processing.

## THE TIME COURSE OF IC

Although no eye-tracking study (that we know of) has tested the lexico-semantic hypothesis specifically, a number of studies have addressed the time course of the use of implicit causality information during processing. There are two competing accounts concerning the time course of the effect of IC bias on pronoun resolution (which are orthogonal to the question of whether IC is the product of lexical semantics or of world knowledge). According to the *focusing* account (e.g., McKoon et al., 1993; Greene and McKoon, 1995), IC information is used immediately after the verb has been encountered, and exerts a top-down influence on pronoun resolution, either by putting one of the verb's arguments in focus or by creating the expectation of a particular kind of explanation. When an anaphor is encountered, the argument that is either in focus or most congruent with the expected explanation is an obvious candidate with which to form a coreference relation.

According to the *integration* account (e.g., Garnham et al., 1996; Stewart et al., 2000), IC information only noticeably starts exerting its influence during processing as soon as it is clear which argument is being rementioned, and (part of) the meaning conveyed by the *because* clause is integrated with the meaning conveyed by the main clause. It may or may not be the case that an expectation about a certain kind of explanation

exists before the pronoun has been encountered, but if it does exist, it only exerts measurable influence on processing as soon as disambiguating information is encountered. The relative slowdown that occurs in explanations that are incongruent with the verb's bias, relative to congruent explanations, is due to the greater difficulty in integrating the meanings of the two clauses.

A series of experiments employing written stories with unambiguous pronouns has provided evidence against the hypothesis that IC bias is only used during sentence wrap-up, when the causal relation between the propositions expressed by the main clause and the *because* clause is presumably checked against world-knowledge (Cozijn et al., 2011b). If the gender of the pronoun matches the gender of only one of the antecedents, readers and listeners can, in principle, ignore IC bias when interpreting the pronoun. Nonetheless, Koornneef and colleagues (Koornneef and Van Berkum, 2006; Koornneef and Sanders, 2013; Koornneef et al., 2015; see also Featherstone and Sturt, 2010) have found that a pronoun which is incongruent in gender with the IC biased anaphor immediately leads to longer reading times, and Van Berkum et al. (2007) have shown that an IC incongruent pronoun elicits a P600 effect compared to an IC congruent pronoun. However, many authors have argued that studies employing unambiguous pronouns cannot distinguish between the focusing account and a version of the integration account in which information derived from the pronoun immediately slows down processing (a.o., Koornneef and Van Berkum, 2006; Kehler et al., 2008; Cozijn et al., 2011a; Koornneef and Sanders, 2013; Järvikivi et al., 2017). Therefore, evidence from different experimental paradigms is needed.

Seven visual world studies on IC have provided such evidence in favor of the focusing account (Cozijn et al., 2011a Experiment 1 and 2; Itzhak and Baum, 2015; Järvikivi et al., 2017 Experiment 1 and 2; Pyykkönen and Järvikivi, 2010; Schlenker and Felser, 2016). In studies like these, participants listen to sentences with IC verbs such as the Dutch example (2), while watching a visual scene that includes the NP<sub>1</sub> and NP<sub>2</sub> referents. Crucially, the pronoun in the *because* clause is ambiguous, so participants can initially only use soft constraints like IC bias to resolve the pronoun. The question is at which time participants start looking more at the referent that is congruent with the IC verb's bias than at the IC-incongruent referent. The auxiliary hypothesis (or *linking hypothesis*) behind this approach is that listeners tend to look at the character or object that they believe is being mentioned at that point in time (or going to be mentioned, as in the seminal study Altmann and Kamide, 1999).

- (2) De octopus<sub>NP1</sub> verveelde<sub>v</sub> de krokodil<sub>NP2</sub> in de auto omdat<sub>CON</sub> hij<sub>PRO</sub>  
*The octopus bored the crocodile in the car because he*  
 rusteloos was en omdat hij het verhaal al  
*restless was and because he the story already*  
 zeker tien keer had verteld<sub>IC-congruent</sub>/ gehoord<sub>IC-incongruent</sub> tijdens de rit.  
*at least ten times had told/heard during the ride.*

“The octopus bored the crocodile in the car because he was restless and because he had already heard/told the story at least ten times during the ride.”

All seven visual world studies on IC report more looks toward the IC-congruent character [the octopus in (2)] than toward the IC-incongruent character [the crocodile in (2)] before the presentation of disambiguating information. The time windows during which IC starts to influence looks range from before the onset of the causal connective, between 1200 and 1500 ms after the onset of the verb (Pyykkönen and Järvikivi, 2010); to between ~400 and 1400 ms after the onset of the connective (Cozijn et al., 2011a, Experiment 1).

Cozijn et al.'s (2011a) two visual world studies employed sentences like (2). In Experiment 1, their target sentences included two consecutive *because* clauses after the main clause. The first *because* clause was globally ambiguous with regard to the referent of the ambiguous pronoun, but the second did ultimately disambiguate the referent. The task was to name the referent of the ambiguous pronoun at the end of the sentence. Participants already showed more looks toward the IC-congruent character than the IC-incongruent character between ~400 and 1400 ms after the onset of the first connective, before the onset of the second connective. In Experiment 2, the globally ambiguous *because* clause was omitted, 60 filler items were included and the task was changed, so that participants only answered comprehension questions about 25% of the items (never about the referent of the ambiguous pronoun). Despite these changes, Experiment 2 showed a positive effect of IC bias on looks toward the IC-congruent character comparable to the effect found in Experiment 1, and during a comparable time window. In sum, both experiments provided evidence in favor of the focusing account and against the integration account.

Pyykkönen and Järvikivi (2010) employed Finnish sentences like (3). The target sentence was preceded by a single sentence that introduced one of the two characters. Besides the type of IC verb (NP<sub>1</sub> or NP<sub>2</sub> biased), Pyykkönen and Järvikivi also manipulated the ambiguous pronoun, so that it was either in nominative or in accusative case. Participants were instructed to look at the screen and listen attentively, while occasionally having to produce an ending to the story they heard. The positive effect of IC bias on looks toward the IC congruent character [the guitarist in (3)] was significant across participants and items from 1200 ms after the onset of the verb, which was some time after the average onset of the object, and remained significant at the onset of the connective and the onset of the pronoun. From the onset of the verb to the disambiguation near the end of the sentence, there was also a strong preference for looking at the subject of the main clause [the butler in (3)] rather than the object.



- (3) Hovimestari<sub>NP1</sub> pelkäsi<sub>v</sub> kitaristia<sub>NP2</sub> ravintolasalissa  
*The butler feared the guitarist in the dining room*  
 koska<sub>CON</sub> koko päivän hän<sub>PRO</sub> kummallista kyllä oli näyttänyt  
*because all day he curiously enough had seemed*  
 erittäin tyytymättömältä huolimatta tulossa olevasta  
*extremely unhappy despite upcoming of the*  
 suosituista illasta.  
*popular night.*

“The butler feared the guitarist in the dining room because for the whole day he curiously enough had seemed extremely unhappy despite the upcoming popular night.”

Like Pyykkönen and Järvikivi (2010), and Järvikivi et al. (2017) conducted their visual world study in Finnish, using sentences that were structurally similar to (3). In Experiment 1, Järvikivi et al. (2017) manipulated IC bias and word order (SVO vs. OVS). Participants already looked more at the IC-congruent referent between 1100 and 1300 ms after the onset of the pronoun. In contrast, the effect of order of mention started between 1300 and 1500 ms. In Experiment 2, Järvikivi et al. manipulated IC bias and the type of pronoun (personal or demonstrative). In this experiment participants already started showing more looks to the IC-congruent referent between 500 and 700 ms after the onset of the pronoun. There were no other main effects, but there were interactions between order-of-mention and type of pronoun and IC congruency and type of pronoun. Importantly, the main effect of IC congruency started before either interaction effect.

In summary, the visual world studies differed in terms of the materials used (the language and the structure of the target sentence), the participants' task and the design of the experiment, but there were also some commonalities. The main clause always consisted of the IC verb with a subject and an object argument, and a filler phrase (usually a prepositional phrase indicating the location of the state or event). Word order was always SVO, except in Järvikivi et al. (2017), Experiment 1, where word order was manipulated. The subordinate clause was always introduced by a causal conjunction, and after the ambiguous pronoun there were one or more filler words that postponed the disambiguation of the referent until later in the sentence. As stated above, all of these studies have found an effect of congruency with IC bias before the presentation of disambiguating information.

Another commonality among the visual world studies on IC is that none of them has addressed the question of whether IC bias is a product of language structural features (the verb's semantic structure and the causal discourse marker) or the product of world knowledge. In order to address this question, it is necessary to manipulate factors of the sentence or context that do not influence the verb's semantic structure or the discourse relation between the clauses. Adding a larger discourse context is ideally suited for this purpose because it allows for the manipulation of characteristics of the referents of the verb's arguments and the situation in which the eventuality takes place without the need to change the sentence in which the IC verb is embedded itself.

The story continuation study reported in Van den Hoven and Ferstl (2018), using stories like those in **Table 1**, has shown that relevant information in the discourse context affects remention patterns. After embedding in a discourse context, remention patterns are still reliably predicted by IC bias, but the predictive

power of IC bias is considerably reduced in a discourse context, and the relevant information in the context also affects the choice of a topic to continue the story. Because of these properties, the stories used in the story completion study provide an excellent way of distinguishing between the biasing characteristics of the verb and the influence of more general knowledge of the world and the discourse context during processing. In the following sections we report two visual world paradigm experiments that were designed to dissociate between the online influence of these two factors on pronoun resolution.

## EXPERIMENT 1

The aim of the current study is to investigate whether the differences in the discourse context which modulate rementions also immediately modulate looks toward the referents (as predicted by the world knowledge account), or whether the effect of discourse context is delayed with respect to the effect of IC bias on looks toward the IC-congruent character (as predicted by the lexical semantics account). In line with previous visual world studies on implicit causality, we expect the IC bias-congruent referent to attract more looks during the *because*-clause than the IC bias-incongruent argument, when the sentence includes a strongly biased verb. That is, if IC Bias influenced viewing patterns we would expect the proportion of looks to NP<sub>1</sub> to be largest for NP<sub>1</sub> verbs, followed by equibiased verbs, and smallest for NP<sub>2</sub> verbs (i.e., a linear decrease). The hypothesis for the discourse context manipulation was that the NP<sub>1</sub> context would increase looks to NP<sub>1</sub>, whereas NP<sub>2</sub> context decreases looks to NP<sub>1</sub>. Finally, an exploratory question was whether such an effect of discourse context would be larger for equibiased verbs compared to more strongly biased verbs.

## Methods

### Participants

Thirty-six participants took part in this experiment. One participant was excluded for not being a native German speaker, and another participant's data were lost due to a technical failure. Among the remaining 34 native speakers of German (12 men, 21 women, 1 queer), age ranged from 19 to 34 ( $M = 24.2$ ,  $SD = 4.4$ ). Vision was normal ( $n = 24$ ) or corrected to normal with contact lenses ( $n = 5$ ) or glasses ( $n = 5$ ). All participants reported having no hearing problems, dyslexia or other potentially relevant bodily or neurological conditions. Participants either received €13 or 1.5 credit hours for compensation.

## Materials

From the 72 story pairs employed by Van den Hoven and Ferstl (2018), we selected the 40 story pairs with the greatest difference in the proportion of NP<sub>1</sub> rementions between the NP<sub>1</sub> biased discourse condition and the NP<sub>2</sub> biased discourse condition (see **Table 1** for an example; see <https://osf.io/9anv2/> for all 72 story pairs, the data and the analysis script for Van den Hoven and Ferstl's (2018) story completion study). The criterion for inclusion was the *difference* in the proportion of NP<sub>1</sub> rementions between story versions; NP<sub>1</sub> biased discourse contexts were often only NP<sub>1</sub> biased relative to the NP<sub>2</sub> biased discourse context. For 30 of the stories, the NP<sub>1</sub> biased story version showed a numerical NP<sub>1</sub> bias, and for 21 stories, the NP<sub>2</sub> biased story version showed a numerical NP<sub>2</sub> bias. Only 11 stories showed a numerical NP<sub>1</sub> bias in the NP<sub>1</sub> biased version and a numerical NP<sub>2</sub> bias in the NP<sub>2</sub> biased version. Implications of this fact are discussed in the section General Discussion.

Each of the stories was constructed around an IC verb. Concerning the verbs' semantic categories, we followed Van den Hoven and Ferstl's (2018) modification of Bott and Solstad's (2014) classification of IC verbs. This classification initially divides verbs into action verbs [agent-patient (*a-p*) verbs] and state verbs on the basis of whether the subject performs an action or experiences or causes a particular emotional state. Among state verbs, stimulus-experiencer (*s-e*) verbs, such as *please*, can be dissociated from experiencer-stimulus (*e-s*) verbs, such as *like* on the basis of argument realization. For instance, the stimulus argument, but not the experiencer, can also be realized by means of a sentential complement.

To this tripartite division, Bott and Solstad (2014) added two classes of verbs which are ambiguous, either because they have an *a-p* and an *s-e* reading (e.g., *frighten*), or because they have an *a-p* reading and an *e-s* reading (e.g., *worship*). Bott and Solstad (2014) also included a class of *a-p* verbs with causal presuppositions, similar to the class of agent-evocator verbs (Au, 1986). Verbs are said to have a causal presupposition when, after the verb is negated, the implication that one of its arguments is causally responsible for the event described by the verb remains, as in *John did not criticize Mary* >> *Mary did something blameworthy*. However, verbs other than unambiguous *a-p* verbs also show this presupposition-like property. For instance, the *e-s* verb *pity* is associated with a causal implication that the object is in an unfortunate situation. Therefore, Van den Hoven and Ferstl (2018) treated causal implications as a dimension orthogonal to verb class, instead of creating additional verb classes.

Following this classification, there were 18 unambiguous *a-p* verbs (e.g., *anrufen/call*); 8 unambiguous *e-s* verbs (e.g., *bemitleiden/pity*); 5 unambiguous *s-e* verbs (e.g., *aufregen/agitate*); and 9 ambiguous *s-e/a-p* verbs (e.g., *beeinflussen/influence*). Twelve of the verbs were associated with a causal implication about the NP<sub>2</sub>. Among these verbs, 10 were unambiguous *a-p* verbs (e.g., *criticize/kritisieren*); 1 was an unambiguous *e-s* verb (*bemitleiden/pity*); and 1 was an ambiguous *s-e/a-p* verb (*stärken/invigorate*). Lemma frequency

of the IC verbs ranged from 0.8 to 127.7 per million in the dlexDB corpus (Heister et al., 2011).

In line with previous visual word studies on IC, we treated IC Bias as a categorical variable. In contrast with previous studies, however, our stories included not only NP<sub>1</sub> biased and NP<sub>2</sub> biased verbs, but also equibiased verbs. The verbs' biases were determined by means of binomial tests on the sentence completion data reported in Van den Hoven and Ferstl (2018), which approximately followed the sentence completion procedure outlined in Garvey and Caramazza (1974). If, among the 34 sentence continuations that were collected for each verb, the NP<sub>1</sub> was rementioned significantly more often than would be expected by chance ( $\alpha = 0.05$  throughout, unless stated otherwise), it was categorized as an NP<sub>1</sub> verb; if the NP<sub>2</sub> was rementioned significantly more often than chance, it was categorized as an NP<sub>2</sub> verb; and if neither was mentioned significantly more often than chance, it was categorized as an equibiased verb. There were 12 NP<sub>1</sub> verbs, 17 NP<sub>2</sub> verbs and 11 equibiased verbs.

On average, 4 words (SD = 1.59) were manipulated between discourse conditions in order to create the expectation of an NP<sub>1</sub> or an NP<sub>2</sub> remention. This was done in various ways, and in five of the stories multiple types of manipulations were employed. Some stories created the impression that agents of *a-p* verbs were insincere (as in **Table 1**;  $n = 9$ ) or misinformed (e.g., blaming the wrong person;  $n = 8$ ), which often led to more explanations about the agent than stories in which the agent was sincere or well-informed, because the internal reasons for the agent's actions are prioritized over the external reasons relating to the patient (Van den Hoven and Ferstl, 2018). (Note, however, that it is possible to highlight an internal reason without rementioning the agent first, as in "John helped Mary because she was, in his eyes, incapable of managing it herself.") Another discourse manipulation was to create the impression that experiencers of *s-e* verbs or *e-s* verbs were unlike most experiencers (e.g., being more naïve or credulous), leading to more rementions of the experiencer ( $n = 25$ ). In these stories, "covariation" information (Kelley, 1967, 1973) was manipulated: There was low consensus between the experiencer and an explicit or implicit reference group on how to evaluate a stimulus, while the stimulus was one of many (potential) stimuli that would evoke the same reaction in the experiencer (i.e., it was not distinctive). Yet another manipulation was to imply that agent/stimulus arguments of ambiguous *s-e/a-p* verbs acted intentionally, leading to more NP<sub>2</sub> rementions, or unintentionally, leading to more NP<sub>1</sub> responses ( $n = 3$ ). This is because when an action is intentional, it is more likely to evoke external reasons, i.e., causes that do not highlight the agent's mind, but which do involve volition on the side of the agent, as in "John amused Mary because she was bored" (Bott and Solstad, 2014). And external reasons usually mention the NP<sub>2</sub> first. Unintentional events, on the other hand, are generally explained by means of direct causes (i.e., causes that do not involve volition on the side of the causer, as in "John amused Mary because he had a funny way of speaking"), which usually remention the NP<sub>1</sub> first. On average, the discourse preceding the target sentence was 3.4 sentences long (SD = 0.7).

In order to make the pronoun in the target sentence ambiguous, the stories were altered to involve two same-gender protagonists (19 stories involved two women and 21 involved two men). In the story completion study, the gender of the protagonists was controlled for by presenting each story version in two gender conditions: male NP<sub>1</sub> and female NP<sub>2</sub> and vice versa. There was no evidence for a gender effect in the materials. It cannot be ruled out that there is an interaction between NP<sub>1</sub> gender and NP<sub>2</sub> gender (for some verbs), such that when both are men or both are women, remention patterns are different from when mixed gender pairs are used. However, there is no evidence for such an effect in the literature, and known gender effects are small and differ between studies (Ferstl et al., 2011; Hartshorne, 2014).

The target sentence always consisted of a main clause with VSO order, followed by a subordinate clause introduced with *because*. In most studies (e.g., Pyykkönen and Järvikivi, 2010; Cozijn et al., 2011a), an auditory distractor region is included between the NP<sub>2</sub> and the pronoun, in order to dissociate more clearly between the effect of the NP<sub>2</sub> and the effect of the pronoun. However, we decided not to include a distractor region, for the following reasons. First of all, there is a trade-off between creating materials that are useful for the eye-tracking record and having the materials sound natural. Every word that is added to the story in order to lengthen the auditory region of interest, or to direct eye-movements to a neutral position, may also be expected by participants to somehow be relevant to the story. This may cause participants to make unintended inferences (e.g., one way to continue the sentence “The butler frightened the guitarist in the dining room” used in Pyykkönen and Järvikivi, 2010, relates to the PP: “because unlike the kitchen it had some great places to hide”). Relatedly, there was no distractor region in the story completion materials, so it was unclear what such a region would do to remention expectations. Moreover, adding words between the conjunction and the pronoun (thus leaving the story as it was in the story completion study up until the conjunction, apart from the character’s genders) is not felicitous in German.

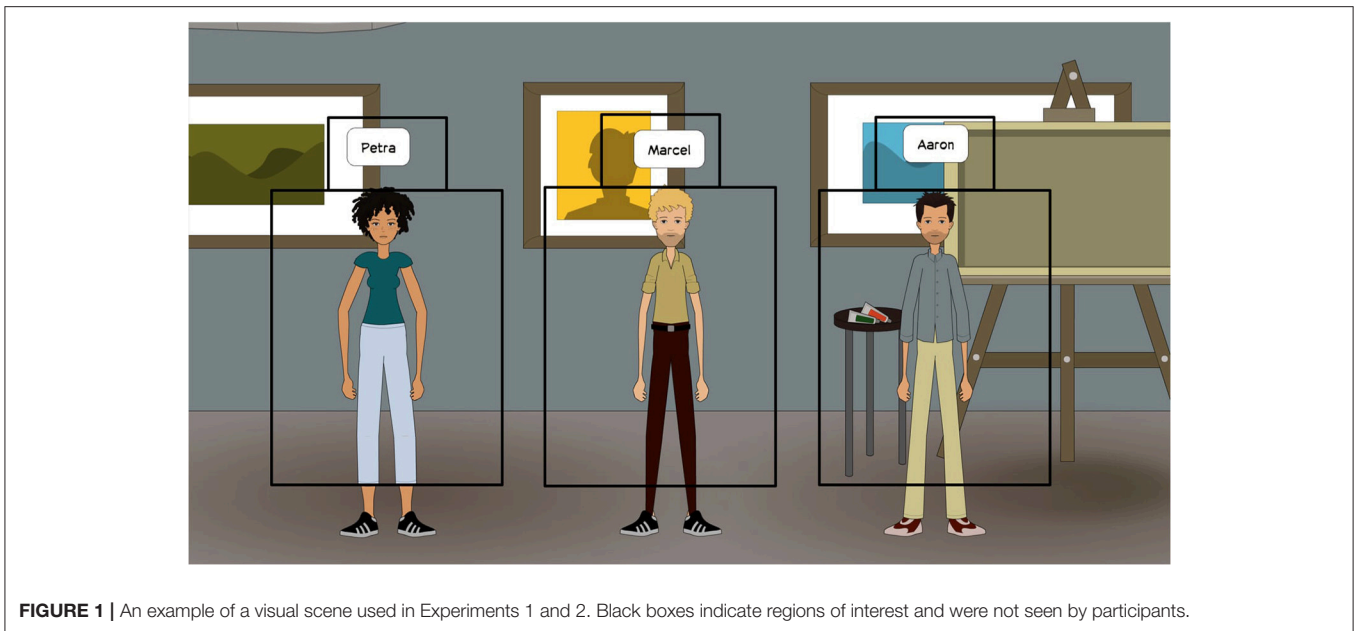
For each story an ending was created, consisting of a story-congruent *because* clause and a concluding sentence. After the word *weil* (*because*), the *because* clause included a filler region of minimally two syllables that was the same for both conditions and that was designed to be minimally informative about who the referent of the ambiguous pronoun might be (e.g., *simply*, *somehow*). After the filler region, the stories in the two discourse conditions diverged in most cases, so that the explanation was always coherent with the rest of the story. The referent of the pronoun, which was no longer ambiguous at the end of the story, was always the same in both discourse conditions. In half of the stories the referent was the character biased by the discourse and in the other half it was the other character. In 14 of the 29 stories that included a non-equibaised verb, the referent was congruent with the bias of the verb and in 15 it was incongruent with the verb’s bias. Among the 11 items with equibaised verbs, 5 continuations rementioned the NP<sub>1</sub> and 6 rementioned the NP<sub>2</sub>. All 40 story pairs used in Experiment 1 can be found in the **Supplementary Materials**, as well as a spreadsheet with descriptive information about every story,

including the remention biases in different story conditions (and remention biases without a story context); information about the semantic category of the verb; and the type(s) of manipulation(s) used.

The two versions of the stories were recorded by a male native German speaker in a quiet room, using a Samson Meteor microphone and the editing software program Reaper v5.15. The same base audio file was used, but for one of the conditions, the sentences that differed from the other condition were replaced by the corresponding sentences from the audio file for the other story version, so that the final audio files only differed at the points where the sentences were not identical between conditions. The same recording of the main clause and the beginning of the subordinate clause of the target sentence was used in both conditions. When the target sentences diverged after the filler region, only the divergent part was substituted in one of the conditions. This was done to ensure that any difference in eye-movements between conditions could not be attributed to differences between the two versions of the target sentence in terms of paralinguistic variables, such as intonation or speech rate. Auditory regions of interest started at the onset of *weil* (*because*) and ended either at the offset of the filler region or after 1000 ms. On average, the auditory region of interest lasted 827 ms (SD = 119, range = 588–1000).

Sixty filler stories were created in order to make it difficult for participants to infer the aim of the study. Thirty-two stories were adaptations of stories used in Van den Hoven and Ferstl’s (2018) story completion study and 28 were newly created stories with a similar structure. Forty-seven fillers had a sentence that resembled the target sentence in the critical stories, but after the main clause these stories continued with a connective that signaled a discourse relation other than an Explanation relation (e.g., Result, Violated Expectation, Denial of Preventer; see Kehler, 2002). This way, participants could not reliably predict an upcoming Explanation relation. In 17 of these fillers, the story continued after the main clause with a reference to a third character which was not mentioned in the main clause. This was done to make the third character potentially relevant at all times. In 15 other stories the NP<sub>1</sub> was mentioned and in another 15 the NP<sub>2</sub> was mentioned. The remaining 13 filler stories did not have a sentence that resembled the target sentence of the critical stories, so that the structure of the stories was somewhat more difficult to predict.

For all 100 stories used in the experiment, a visual scene was created using the online comic creation tool Pixton ([www.pixton.com](http://www.pixton.com); see **Figure 1** for a version of the visual scene corresponding to the story in **Table 1**). The visual scene depicted the two characters referred to in the main clause, as well as a third character who was mentioned in some, but not all of the stories. The characters were embedded in an environment that was congruent with the content of the story, but they always had the same neutral posture and facial expression. The position of the characters differed per item, with each of the characters (NP<sub>1</sub>, NP<sub>2</sub>, and other) occurring approximately equally often in all three positions (each character was in each position between 29 and 39 times). For target items, two versions of the same visual scene were created, with the NP<sub>1</sub> and NP<sub>2</sub> characters in swapped



**FIGURE 1** | An example of a visual scene used in Experiments 1 and 2. Black boxes indicate regions of interest and were not seen by participants.

positions (but the background remained unaltered). During the story presentation, the characters' names were presented above the characters for identification. Separate regions of interest were specified for the characters and their names, but in the analysis looks toward the name and looks toward the character were not differentiated, because our hypothesis did not specify any differences in looks to the names or the characters.

For 90 stories, a comprehension question was created. Forty questions directly queried whether the participants could identify the characters by their names, and fifty focused on a fact about the story. From the remaining ten stories (all of which were filler items), the final sentence was omitted and participants were asked to formulate an ending to the story, which was recorded.

We created four experimental lists of 100 stories, in which Discourse Context and the positions of the NP<sub>1</sub> and NP<sub>2</sub> characters were counterbalanced in a Latin square design. The stories were presented in a pseudo-randomized order which was different for every participant. Participants never saw more than four consecutive target items.

## Procedure

According to the guidelines of the German Research Foundation (DFG, see [http://www.dfg.de/foerderung/faq/geistes\\_sozialwissenschaften/index.html](http://www.dfg.de/foerderung/faq/geistes_sozialwissenschaften/index.html)), the present study was exempt from ethics approval, because all participants were adult volunteers who were fully informed, the study did not involve risks, and it did not pose undue physical or emotional demands. In the instructions, participants were requested to listen attentively to the stories so that they would be able to answer a short comprehension question after every story. During the stories, participants did not have an explicit task apart from to "look and listen" (Kamide et al., 2003). The experiment started with five practice items, one of which required participants to formulate an ending to the story. A short break was included

after approximately every 25 stories. The experiment took around 1.5 h.

Eye-movements were recorded with a monocular SR Research EyeLink1000 eye tracker, at a sampling rate of 1000 Hz. The images were presented with a resolution of 1600 by 1200 pixels at a display rate of 60 Hz on a monitor that was located 60 cm from the participants' eyes. A chin rest was used for head stabilization. Nine-point calibration was performed before the first trial and after every break. Every trial started with a drift check. The stories were presented through headphones.

## Data Analysis

Trials with eye-tracker loss (due to blinks, looks beside the monitor, etc.) during more than 50% of the recorded frames in the relevant time window were excluded (55 trials; 4% of the total number of data points).

The eye-movement data were analyzed using a non-parametric permutation test involving the comparison between a linear mixed effects model fitted to the data and reference linear mixed effects models fitted to permuted datasets, which allowed us to test whether our predictors were associated with an increase in looks to the NP<sub>1</sub> character at any time during our auditory region of interest, without the need for specific hypotheses about when looking patterns would diverge (Maris and Oostenveld, 2007; Wittenberg et al., 2017).

The data were divided into 100 ms bins. The dependent variable was binary: 1 if the participant looked toward the NP<sub>1</sub> more than toward the NP<sub>2</sub> across all milliseconds without tracker loss in a given time bin, and 0 if the participant looked toward the NP<sub>2</sub> more than toward the NP<sub>1</sub>. Time bins in which neither the NP<sub>1</sub> nor the NP<sub>2</sub> was fixated were excluded from the analysis (13.4 % of the total number of data points). The independent variables were IC Bias and Discourse Context. For IC Bias, two orthogonal contrasts were used (Field et al., 2012). The first



contrast (IC Bias: 1) tested the hypothesis that the proportion of looks to the NP<sub>1</sub> is largest for NP<sub>1</sub> verbs, followed by equibaised verbs, and smallest for NP<sub>2</sub> verbs (i.e., a linear decrease). The second contrast code, consequently, compared equibaised verbs to the two strongly biased verb categories (NP<sub>1</sub> and NP<sub>2</sub>). Discourse Context was a categorical variable with two levels: NP<sub>1</sub> biased and NP<sub>2</sub> biased. For Discourse Context, effect coding was used (NP<sub>2</sub> was coded as  $-1$  and NP<sub>1</sub> was coded as  $1$ ).

For each time bin, we used the `glmer()` function from the `lme4` package (Bates et al., 2015a) in R to calculate a mixed-effects logistic regression model with the fixed effect terms Discourse Context, IC Bias and the interaction Discourse Context  $\times$  IC Bias, random intercepts per participant and per story, and random slopes for Discourse Context per participant. Principal Components Analyses on the random effects structure showed that the data did not support a more elaborate random effects structure in all time bins: The number of dimensions included was sufficient to account for 100% of the variance explained (Bates et al., 2015b). Since the random slopes for IC Bias and the interaction Discourse Context  $\times$  IC Bias per participant were the smallest variance components, these were removed from the model specification. For each fixed effects term, clusters were identified as temporally adjacent time bins in which the Wald  $z$  statistic for the term exceeded the thresholds of  $1.6$  or  $-1.6$ . This moderate-sized threshold allowed for the detection of shallow but long-lasting effects without affecting the false alarm rate. Lower values of the threshold lead to more clusters being detected, but the probability of observing a statistically significant cluster statistic does not change. Cluster statistics were calculated by summing the  $z$ -values within each cluster (see Maris and Oostenveld, 2007). Summing the  $z$ -values across time bins leads to a statistic that incorporates both the size of the effect and its duration, and is thus more fine-grained than merely using the number of time bins in which the threshold is exceeded as a test statistic.

To create the empirical distribution against which the cluster statistics were to be compared, we randomly permuted the values of the factors Discourse Context and IC Bias within subjects (in order to control for variability between individuals), while keeping both the time bin and the proportion of trials belonging to each factor level constant. With these permuted predictors, we fit a mixed-effects logistic regression model to the eye-tracking data, and subsequently identified the clusters as described above. If multiple clusters were identified for a given term, only the cluster with the maximum cluster statistic (or minimum cluster statistic for clusters with a negative sign) was included in the empirical distribution (In the data, however, multiple clusters could be detected for each term—each of which would be compared to the maximum cluster statistics from the empirical distribution.). This process was repeated 1,000 times to give an estimate of the probability of obtaining the cluster statistics as large as or larger (or smaller for clusters with a negative sign) than those we obtained given randomly distributed predictor values.

## Results

The mean proportion of correctly answered comprehension questions was  $0.96$  ( $SD = 0.03$ ). Likelihood ratio tests showed

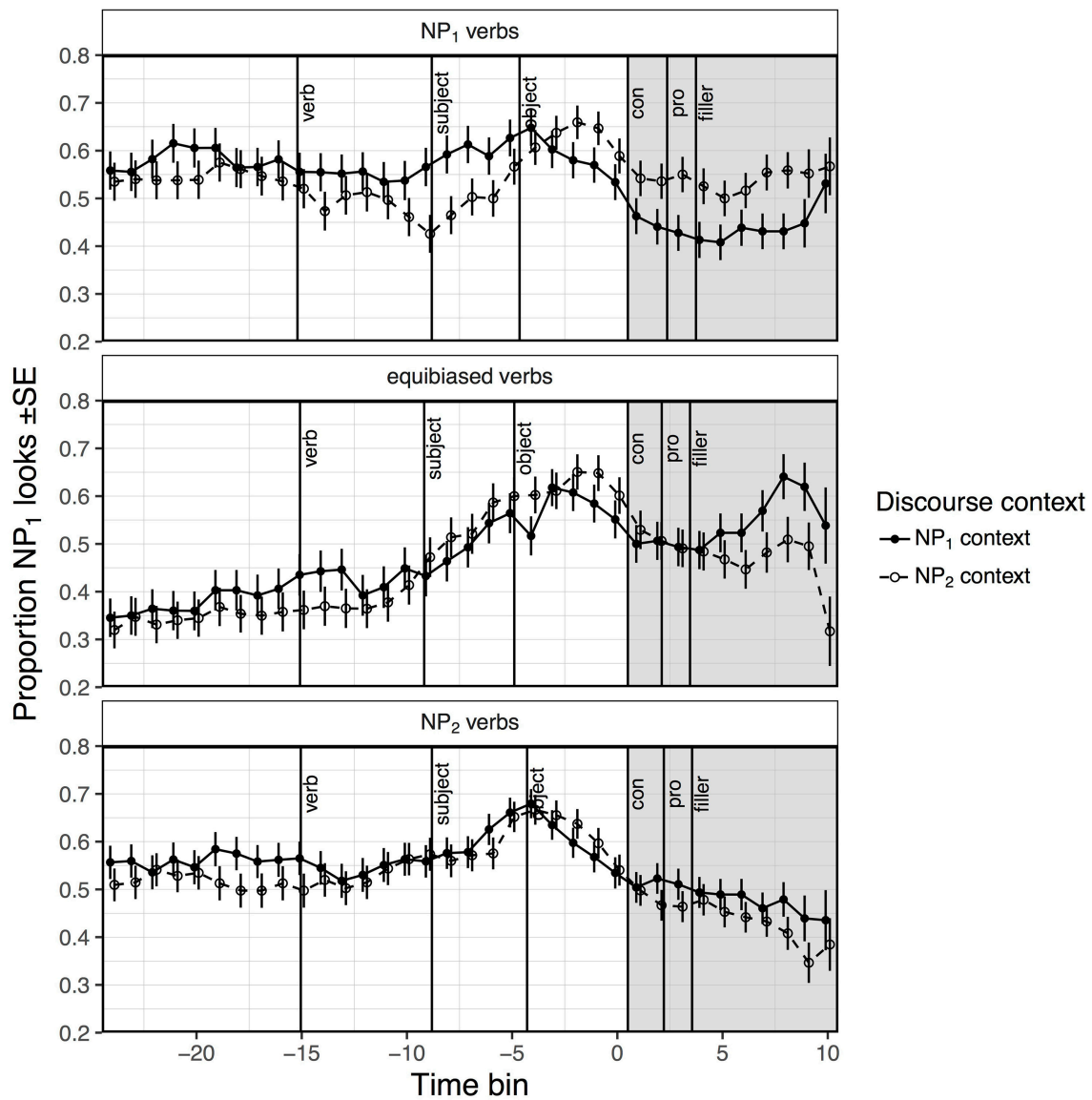
that the model fit to the eye-tracking data which included the interaction Discourse Context  $\times$  IC Bias had a better fit than the model without the interaction term for some of the time bins. **Figure 2** shows the proportion of looks toward the NP<sub>1</sub> character across time bins and across conditions. **Figure 3** shows the  $z$ -values of each term in the model and the empirical distribution of  $z$ -values for the permuted predictors across time bins.

Four clusters were identified: one for the main effect of IC Bias: 1,  $p = 0.121$ ; one for the main effect of IC Bias: 2,  $p = 0.091$ ; one for the interaction Discourse Context  $\times$  IC Bias: 1,  $p = 0.003$ ; and one for the interaction Discourse Context  $\times$  IC Bias: 2,  $p = 0.091$ . **Table 2** shows the cluster statistics, their associated time bins and the proportion of cluster statistics in the empirical distribution that was equal to or greater than the cluster statistic (the  $p$ -value).

To interpret the interaction Discourse Context  $\times$  IC Bias: 1, we performed *post-hoc* analyses of simple effects by restricting the data used in the analysis of one variable to one of the levels of the other variable, and fitting a linear mixed effects model with only the former variable as a fixed effect to the remaining data. The random effects structure and the analysis procedure were the same as for the main analysis, except that the time window was restricted to the window where the cluster for the interaction Discourse Context  $\times$  IC Bias: 1 was found. The *post-hoc* tests revealed that, in the NP<sub>2</sub> biased Discourse Condition, there was a linear decrease in looks to the NP<sub>1</sub> from items with NP<sub>1</sub> biased verbs through equibaised verbs to NP<sub>2</sub> biased verbs during the time window 600–900 ms after the onset of the connective, cluster statistic =  $-6.61$ ,  $p = 0.026$ . No clusters were found for the effect of IC Bias: 1 among items in the NP<sub>1</sub> biased Discourse Condition. Among items with NP<sub>1</sub> verbs, an NP<sub>1</sub> Discourse Context led to fewer looks toward the NP<sub>1</sub> than an NP<sub>2</sub> Discourse Context during the time window 100–500 ms after the onset of the connective, cluster statistic =  $-7.43$ ,  $p = 0.013$ . Among items with equibaised verbs, a non-significant cluster was found for the effect of Discourse Context during the time window 700–800 ms after the onset of the connective, cluster statistic =  $1.80$ ,  $p = 0.871$ . No clusters were found for the effect of Discourse Context among NP<sub>2</sub> biased verbs.

## Discussion

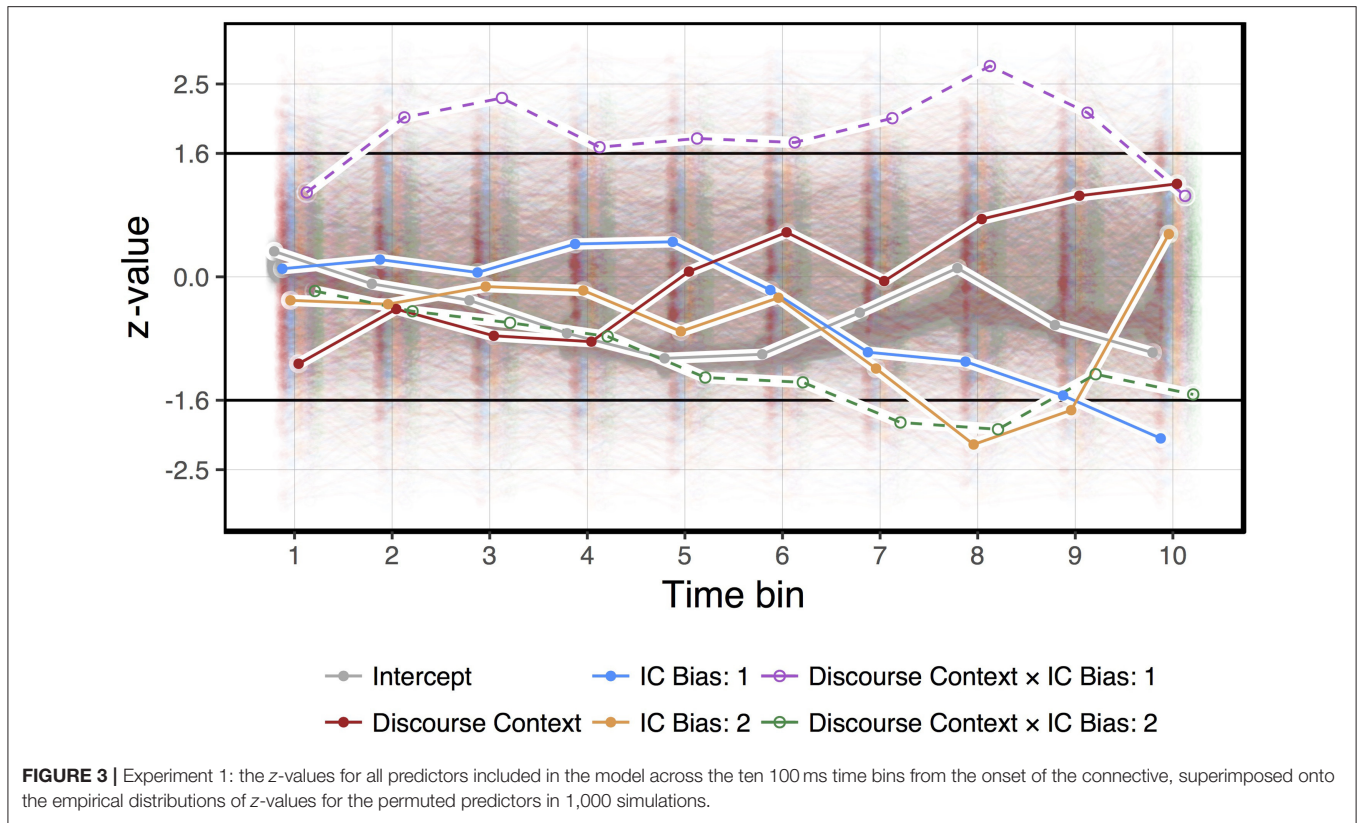
Experiment 1 provided some evidence that listeners use IC bias during pronoun resolution even when the sentence is embedded in a larger discourse context, although this effect only held for verbs embedded in an NP<sub>2</sub> biased Discourse Condition. We found no compelling evidence in favor of the hypothesis that listeners are generally sensitive to the discourse context when resolving ambiguous pronouns in the same way that participants writing a continuation to the story are sensitive to the discourse context when choosing a topic with which to continue the story. Instead of looking at the NP<sub>1</sub> character more often in an NP<sub>1</sub> biased discourse context, participants looked at the NP<sub>1</sub> character *less* often in such a context, and only during items with NP<sub>1</sub> verbs (see **Figure 2**). These results are more in line with an account of IC in which listeners initially only draw upon lexical semantics than with an account in which world knowledge immediately guides pronoun interpretation.



**FIGURE 2 |** The proportion of looks toward the NP<sub>1</sub> character out of looks toward the NP<sub>1</sub> and NP<sub>2</sub> characters per time bin in Experiment 1, by IC Bias and Discourse Context. The time window shows looks from 2500 ms before the onset of the connective until the onset of disambiguating information. Each time bin represents 100 ms. Vertical lines indicate the mean onset of the verb, subject, object, ambiguous pronoun (pro) and filler region per IC bias bin, and the onset of the connective (con), to which the data were time-locked. The shaded area indicates the region that was included in the statistical analysis.

There are several reasons why it is surprising that IC bias showed a simple main effect while discourse context did not reliably affect looks in the predicted direction for any subset of verbs. Throughout the experiment, IC bias was neither a strongly reliable cue for the type of discourse relation that would follow (only 40 out of 87 interpersonal verbs were followed by an Explanation), nor a strongly reliable cue for which character would be mentioned next (in only 18 out of 40 stories with an Explanation relation did the ambiguous pronoun refer to the numerically IC biased character). In natural discourse, IC bias is a strong cue for both of these outcomes (Long and De Ley, 2000), so participants could have adapted to the experimental

environment and stopped relying on IC bias. Moreover, the story pairs were selected because the discourse context had a strong effect on rementions in these stories; the IC verbs did not have particularly strong IC biases in comparison to other visual world studies on IC. Cozijn et al. (2011a), for instance, used NP<sub>1</sub> verbs with an average NP<sub>1</sub> remention bias of 93.4% (SD = 4.1), and NP<sub>2</sub> verbs with an average NP<sub>2</sub> remention bias of 80.1% (SD = 7.3). The remention biases for our NP<sub>1</sub> and NP<sub>2</sub> verbs were only 80.9% (SD = 6.8) and 76.3% (SD = 18.7), respectively, and unlike any previous visual world study on IC, our study included equibaised verbs (11 of our 40 verbs).



**TABLE 2** | Experiment 1: cluster statistics for the model terms, the time window with which the cluster was associated and the proportion of the models that formed the empirical distribution which had a cluster statistic equal to or higher than (or lower than, if the cluster statistic had a negative sign) the cluster statistic found in the data for a given term (the *p*-value).

| Model term                     | Time window (ms) | Cluster statistic | <i>p</i> |
|--------------------------------|------------------|-------------------|----------|
| IC Bias: 1                     | 900–1000         | –2.09             | 0.121    |
| IC Bias: 2                     | 700–900          | –3.90             | 0.091    |
| IC Bias: 1 × Discourse context | 100–900          | 16.52             | 0.003    |
| IC Bias: 2 × Discourse context | 600–800          | –3.86             | 0.091    |

One possible reason for the lack of the hypothesized effect of discourse context is that the rich discourse context may have led to a variety of inferences other than the intended ones, attenuating the effect. Relatedly, the number of stories and their length might have made it difficult for participants to sustain attention. However, there are two reasons why a lack of attention is an unlikely cause for the pattern of results. First, performance on comprehension questions was almost at ceiling. And second, as **Figure 2** shows, participants did look more at the NP<sub>1</sub> rapidly after the onset of the subject in the main clause, and less at the NP<sub>1</sub> sometime after the onset of the object. The only condition in which participants did not directly look less at the NP<sub>1</sub> after the onset of the object, was the NP<sub>1</sub> biased discourse for NP<sub>1</sub> verbs.

A final factor that may have caused noise in the data is that the German sentences included an unintended ambiguity. The main clause “...enttäuschte Colin Steffen...” (“...disappointed Colin Steffen...”) can either be interpreted as having the intended, canonical VSO order, or as having a scrambled VOS order. The VOS order is dispreferred for all verbs, but it is less dispreferred for stimulus-experiencer verbs than for experiencer-stimulus verbs (Scheepers et al., 2000) and presumably for action verbs. Our stimuli included 12 verbs that can be interpreted as stimulus-experiencer verbs, and they are mostly NP<sub>1</sub> biased, so subject-object ambiguity might have led to noise in looking patterns that could hide the effect of discourse context. Four of the comprehension questions directly queried the interpretation of stimulus-experiencer verbs: *Who discouraged the other?*; *Who disturbed the other?*; *Who amused the other?*; and *Who was shocked?* Out of all 136 answers to these four questions, only 2 were incorrect, so participants nearly always arrived at the intended interpretation by the end of the story, but the possibility that they initially had a different interpretation in mind, cannot be excluded.

## EXPERIMENT 2

In order to rule out several of the alternative explanations why discourse context failed to show the hypothesized effect, we carried out a second visual world experiment, in which the word order of the main clause was unambiguous. The hypotheses were the same as for Experiment 1. According to both the lexical

semantics account and the world knowledge account, NP<sub>1</sub> biased verbs should lead to more looks to the NP<sub>1</sub> character than NP<sub>2</sub> biased verbs, but only the world knowledge account predicts that an NP<sub>1</sub> biased discourse context immediately leads to more looks to the NP<sub>1</sub> character than an NP<sub>2</sub> biased context.

## Methods

### Participants

Thirty-eight participants were recruited who did not take part in Experiment 1. One participant's data were lost due to a technical failure, and another participant was excluded because of poor quality eye-movement data. Thirty-six native speakers of German remained (11 men, 25 women). Age ranged from 19 to 36 ( $M = 23.9$ ,  $SD = 3.1$ ). Vision was normal ( $n = 27$ ) or corrected to normal with contact lenses ( $n = 5$ ) or glasses ( $n = 4$ ). All participants reported having no hearing problems, dyslexia or other potentially relevant bodily or neurological conditions, except for one, who reported having a potentially relevant condition but did not specify this further. Participants either received €8.50 or 1 credit hour as compensation.

### Materials

The 40 target stories used in Experiment 2 were altered so that the main clause was no longer ambiguous between a VSO reading and a VOS reading. In the adapted stories, the character's names were preceded by an article indicating case, a feature that is common in various Alemannic dialects (the dialects spoken in the area around Freiburg) and other Southern German dialects. Although this feature is marked in the standard dialect, it is understandable to all native speakers of German. Instead of *...enttäuschte Colin Steffen...* (*...disappointed Colin Steffen...*), the main clause was now *...enttäuschte der Colin den Steffen...* (*...disappointed the<sub>nom</sub> Colin the<sub>acc</sub> Steffen...*). Since the feminine definite articles indicating the nominative and accusative case are syncretic (*die* is used for both), and most of our IC verbs had an object in the accusative case, female protagonists were exchanged for male protagonists in both the stories and the images.

Furthermore, in order to rule out an influence of the words in the filler region on the resolution of the pronoun, and to prolong the time window during which the pronoun was ambiguous, we replaced the filler words with 1000 ms of white noise. To limit the duration of the experiment, and in order not to induce a bias against Explanation relations following interpersonal verbs, we included only 24 filler stories. All fillers contained a 1000 ms stretch of white noise at various points during the stories.

Finally, in three experimental stories one of the sentences before the target sentence was changed to make the referential form of the NPs in the main clause sound more natural given the information structure of the discourse ( $n = 2$ ) or to even out the number of references to the characters in both versions of the story ( $n = 1$ ).

### Procedure

Experimental procedures were the same as in Experiment 1, except that participants were told that the audio quality of the stories was not always optimal, but that they should nonetheless

try to understand the stories as well as they could, and ignore the disruptions. Only one break was included after approximately half of the stories. The experiment took ~50 min in total.

### Data Analysis

Trials with eye-tracker loss (due to saccades, blinks, etc.) during more than 50% of the recorded frames in the relevant time window were excluded (102 trials; 7% of the total number of data points). Moreover, time bins during which neither the NP<sub>1</sub> nor the NP<sub>2</sub> was fixated were also excluded (16.6% of the total number of data points). The same analysis procedure was used as for Experiment 1.

### Results

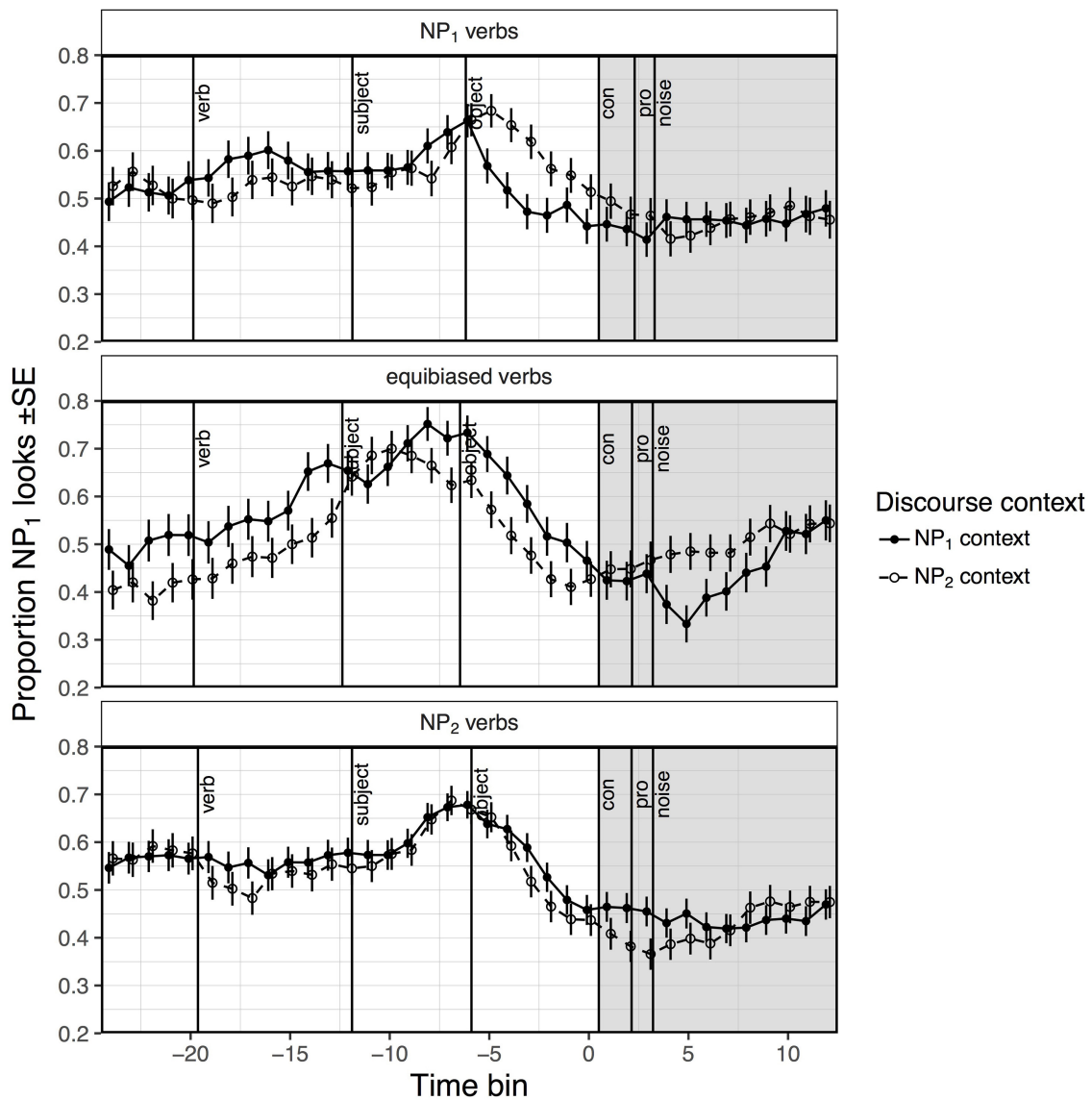
As in Experiment 1, the mean proportion of correctly answered comprehension questions was 0.96 ( $SD = 0.03$ ). Likelihood ratio tests showed that the model with the interaction Discourse Context  $\times$  IC Bias had a better fit than the model without the interaction term for some of the time bins. **Figure 4** shows the proportion of looks toward the NP<sub>1</sub> character across time bins and across conditions. **Figure 5** shows the  $z$ -values of each term in the model and the empirical distribution of  $z$ -values for the permuted predictors across time bins.

Four clusters were identified: one for the main effect of IC Bias: 2,  $p = 0.116$ ; one for the main effect of Discourse Context: NP<sub>1</sub>,  $p = 0.172$ ; one for the interaction IC Bias: 1  $\times$  Discourse Context,  $p = 0.072$ ; and one for the interaction IC Bias: 2  $\times$  Discourse Context,  $p = 0.100$ . **Table 3** shows the cluster statistics, their associated time bins and the proportion of cluster statistics in the empirical distribution that was equal to or greater than the cluster statistic. IC Bias: 1 showed a marginally significant interaction with Discourse Context: moving from items with NP<sub>1</sub> biased verbs through items with equibiased verbs to items with NP<sub>2</sub> biased verbs, looks toward the NP<sub>1</sub> slightly decreased if the Discourse Context was NP<sub>2</sub> biased, but not if the Discourse Context was NP<sub>1</sub> biased. Finally, the interaction Discourse Context: NP<sub>1</sub>  $\times$  IC Bias: 2 was also a marginally significant predictor of looks toward the NP<sub>1</sub>. Among items with NP<sub>2</sub> verbs (and to a lesser degree items with NP<sub>1</sub> verbs), an NP<sub>1</sub> Discourse Context led to more looks toward the NP<sub>1</sub> than an NP<sub>2</sub> Discourse Context, whereas among items with equibiased verbs, an NP<sub>1</sub> Discourse Context led to fewer looks toward the NP<sub>1</sub> than an NP<sub>2</sub> Discourse Context during the time window 300–500 ms after the onset of the connective.

### Discussion

In Experiment 2, NP<sub>1</sub> verbs again seemed to lead to more looks toward the NP<sub>1</sub> than NP<sub>2</sub> verbs during the first 300 ms after the onset of the connective in an NP<sub>2</sub> biased Discourse Context. However, the effect is not straightforwardly interpretable because, as can be seen in **Figure 4**, it starts before the onset of the connective. Moreover, the interaction effect is only marginally significant. Although participants' performance on the comprehension questions was at ceiling, they may not have devoted sufficient attention to the stories to show a rapid influence of IC bias on eye-movements. As in Experiment 1, there



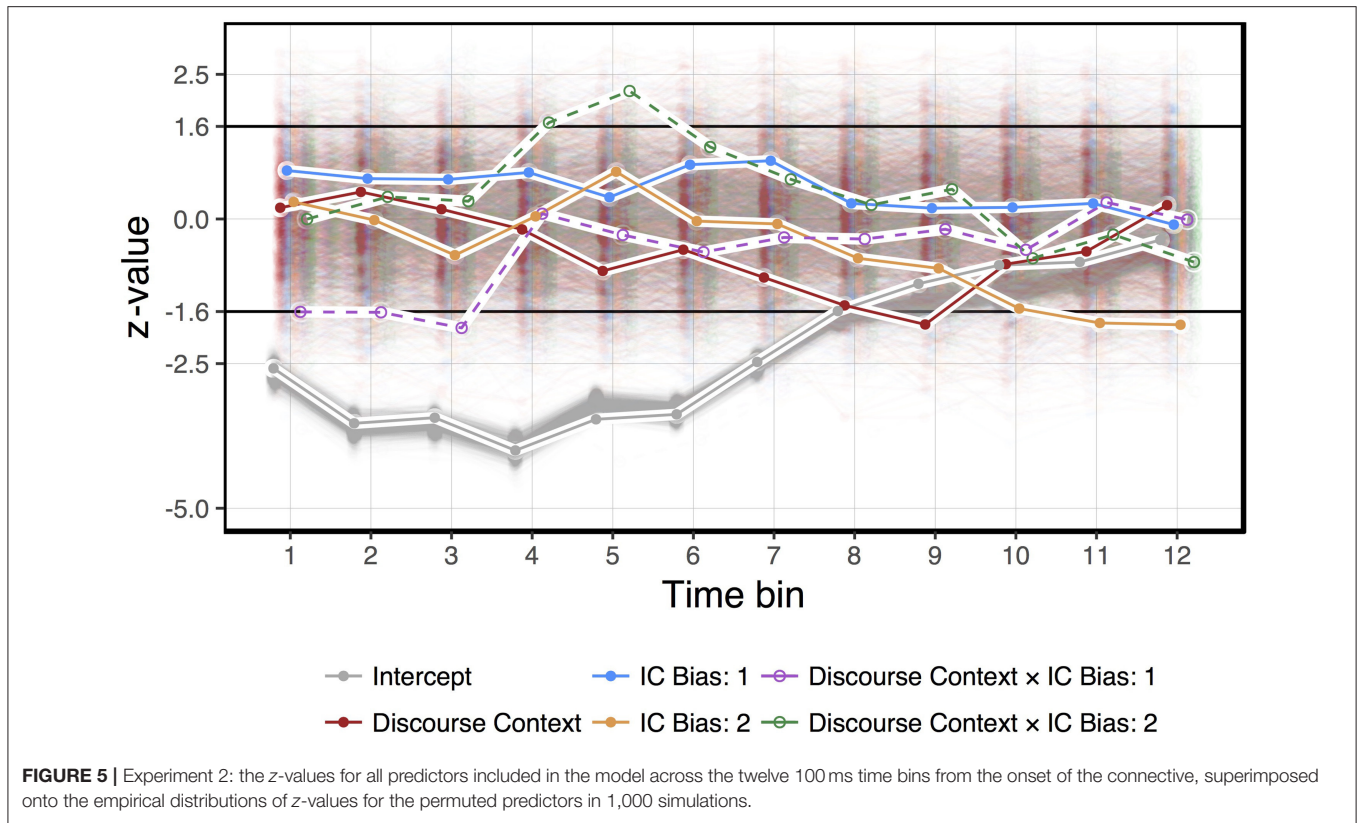


**FIGURE 4 |** The proportion of looks toward the NP<sub>1</sub> character out of looks toward the NP<sub>1</sub> and NP<sub>2</sub> characters per time bin in Experiment 2, by IC Bias and Discourse Context. The time window shows looks from 2500 ms before the onset of the connective until the onset of disambiguating information. Each time bin represents 100 ms. Vertical lines indicate the mean onset of the verb, subject, object, ambiguous pronoun (pro), and filler region per IC bias bin, and the onset of the connective (con), to which the data were time-locked. The shaded area indicates the region that was included in the statistical analysis.

was no main effect of discourse context, but discourse context did interact with IC bias. However, the interaction effects were different in Experiment 2. In Experiment 1, an NP<sub>1</sub> discourse context led to fewer looks toward the NP<sub>1</sub> among items with NP<sub>1</sub> verbs compared to items with NP<sub>2</sub> biased verbs from 100 to 900 ms after the onset of the connective. In Experiment 2 the same interaction occurred (although this time it was only marginally significant) during the first 300 ms after the connective, but from 300 to 500 ms, an NP<sub>1</sub> Discourse Context led to marginally significantly fewer looks toward the NP<sub>1</sub> among items with *equibaised* verbs, compared to NP<sub>1</sub> biased and NP<sub>2</sub> biased verbs.

We again found no evidence in favor of the hypothesis that listeners are immediately sensitive to the discourse context when resolving ambiguous pronouns in the same way that participants writing a continuation to the story are sensitive to the discourse context when choosing a topic with which to continue the story. Participants only looked at the NP<sub>1</sub> character more often in an NP<sub>1</sub> biased discourse context in a subset of the stories, particularly those with NP<sub>2</sub> verbs.

Interestingly, across all three categories of IC verbs, an NP<sub>1</sub> biased context did lead to more looks toward the NP<sub>1</sub> than an NP<sub>2</sub> biased context between the onset of the verb and the onset of the subject, as can be seen in **Figure 4**. This suggests that



**TABLE 3 |** Experiment 2: cluster statistics for the model terms, the time window with which the cluster was associated and the proportion of the models that formed the empirical distribution which had a cluster statistic equal to or higher than (or lower than, if the cluster statistic had a negative sign) the cluster statistic found in the data for a given term (the *p*-value).

| Model term                     | Time window (ms) | Cluster statistic | <i>p</i> |
|--------------------------------|------------------|-------------------|----------|
| Discourse context              | 800–900          | –1.82             | 0.172    |
| IC Bias: 2                     | 1000–1200        | –3.63             | 0.116    |
| IC Bias: 1 × Discourse context | 0–300            | –5.10             | 0.072    |
| IC Bias: 2 × Discourse context | 300–500          | 3.88              | 0.100    |

at this point the discourse context did draw the participants’ attention to the referent that was most likely to be rementioned in an explanation for the event that was being described. Although we found no evidence that this possible attentional bias during the beginning of the main clause also influenced pronoun resolution in the *because* clause, we performed an exploratory analysis to further investigate this potential early attentional bias. Additionally, to investigate what it is about the verb’s lexical semantics that drives the looks toward the IC congruent character during the subordinate clause, we explored whether the eye-movement record showed a distinct pattern for verbs with a more causal NP<sub>1</sub> argument and verbs with a less causal NP<sub>1</sub> argument. The possibility that the causal potency of thematic roles plays an important part in IC has

been explored in an offline study by Hartshorne and Snedeker (2013).

### Exploratory Analysis

We performed two data-driven analyses on the combined data from Experiments 1 and 2. The two analyses had different aims. The aim of the first analysis was to find out whether the eye-tracking record might, at some point during the target sentence, reveal a bias in the attention of the listeners in favor of the referent who was most likely to be rementioned in an explanation. In this analysis, we used the eye-tracking records to predict the remention biases obtained in the story continuation study by Van den Hoven and Ferstl (2018). Remention biases were calculated as the log-odds of rementioning the NP<sub>1</sub> (or NP<sub>2</sub>) in the story continuation study. Since the IC verbs were embedded in a discourse context, the analysis did not differentiate between the effects of IC bias and discourse context. The aim of the second analysis was to find out whether the eye-tracking record would reveal an early bias toward the NP<sub>1</sub> character in the attention of listeners when the NP<sub>1</sub> character had a more causal thematic role (agent or stimulus; e.g., John in *John shocked Mary*) rather than a less causal thematic role (experiencer; e.g., John in *John noticed Mary*) (see Croft, 2012).

We time-locked the eye-tracking record to four different time points: the onset of the verb; the subject (including the article in Experiment 2); the object (*idem*) and the connective. This led to 4 time windows, which were limited to 1500 ms

each. In the first analysis we performed a Student's *t*-test for each millisecond *t* in each time window, in which the outcome variable was the NP<sub>1</sub> remention bias and the predictor was the fixated referent: NP<sub>1</sub> or not-NP<sub>1</sub>. This way, we could ascertain whether the trials in which listeners fixated the NP<sub>1</sub> at a given time *t* were the trials with stories that led to many NP<sub>1</sub> rementions in the story completion task. In the second analysis, we performed a logistic regression analysis for each millisecond *t* in each time window, in which the outcome variable was causal NP<sub>1</sub> argument (yes or no) and the predictor was the fixated referent: NP<sub>1</sub> or not-NP<sub>1</sub>. There were 32 stories with a causal NP<sub>1</sub> argument and 8 stories with a non-causal NP<sub>1</sub> argument. There was a moderate positive point-biserial correlation between remention bias and causal NP<sub>1</sub> argument among the stories,  $r = 0.39$ ,  $p < 0.001$ . The means (for the first analysis), log-odds ratios (for the second analysis), CIs and *p*-values were smoothed by means of a moving window average, using a Hamming window of 101 ms. The same analyses were performed for NP<sub>2</sub> fixations predicting NP<sub>2</sub> rementions and causal NP<sub>1</sub> arguments. We corrected for multiple comparisons using false discovery rate correction (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001), which controls for multiple testing under dependency, and report results for analyses in which  $\alpha = 0.05$  and  $\alpha = 0.005$ .

**Figure 6** shows the difference in mean NP<sub>1</sub> remention bias between trials in which the NP<sub>1</sub> was fixated at time *t* and trials in which the NP<sub>1</sub> was not fixated at time *t*, and the same for NP<sub>2</sub> remention biases and fixations. At 650 ms after the onset of the verb, the trials in which the NP<sub>1</sub> was fixated were trials with stories in which the NP<sub>1</sub> was rementioned on average 53.2% of the time in the story completion study, compared to 49.6% of the time for trials in which the NP<sub>1</sub> was not fixated at that time point.

We found the same pattern around the average onset of the filler, for both NP<sub>1</sub> and NP<sub>2</sub> remention biases and fixations, continuing into the disambiguating region. The fact that the effect holds up to the disambiguating region might be somewhat surprising, given that bias-congruent and bias-incongruent explanations were approximately counterbalanced across conditions, but it should be noted that the onset of the "disambiguating region" did not immediately disambiguate the ambiguous pronoun in all cases. In 8 stories, the non-referent was rementioned in the explanation, as in (4a), but in the 32 other stories, the referent had to be inferred, as in (4b).

(4a) Heimlich hasste Vincent Theo, weil er einfach fand,  
*Secretly hated Vincent Theo because he simply found*  
 dass Theo zu egoistisch war.  
*that Theo too selfish was.*

"Secretly Vincent hated Theo because he simply thought that Theo was too selfish."

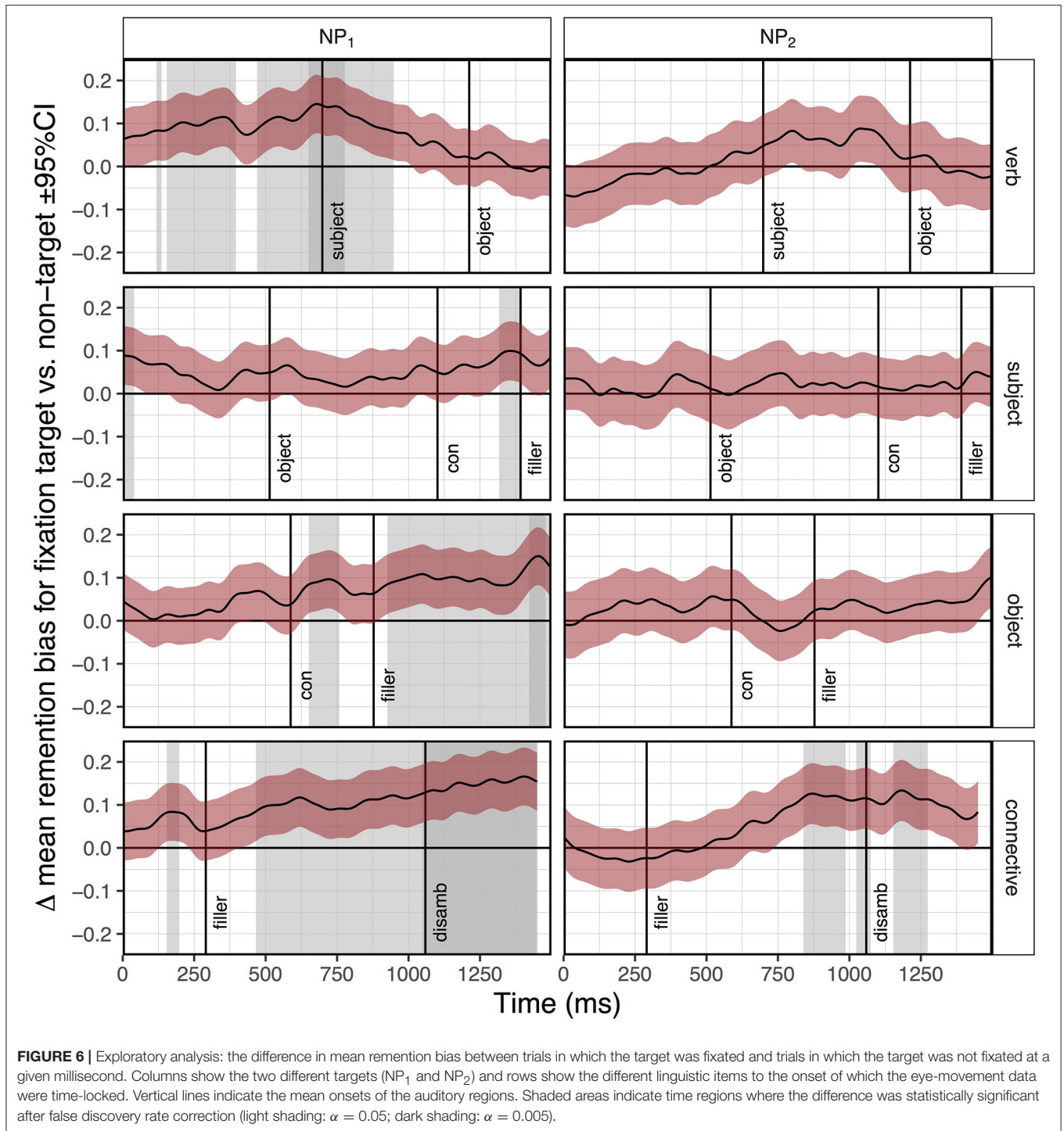
(4b) Im Atelier kritisierte Marcel Aaron, weil er einfach  
*In the workplace criticized Marcel Aaron because he simply*  
 fand, dass die Bilder unprofessionell waren.  
*found that the paintings unprofessional were.*

"In the workplace Marcel criticized Aaron because he simply thought that the paintings were unprofessional."

The point at which information becomes available that can be used to select the most probable referent for the ambiguous pronoun (Cozijn et al., 2011b) differed across stories (and the point at which listeners used this available information is also likely to differ), but was always later than the onset of the disambiguating region. So the relative preference in NP<sub>1</sub> looks around the onset of the disambiguating region is likely due to the use of the remention bias before disambiguation had taken place.

**Figure 7** shows the log-odds ratio of causal and non-causal NP<sub>1</sub> arguments between trials in which the NP<sub>1</sub> was fixated at time *t* and trials in which the NP<sub>1</sub> was not fixated at time *t*, and the same for NP<sub>2</sub> fixations. Unlike in the first analysis, the eye-movements did not predict whether the NP<sub>1</sub> argument is causal or not within 1000 ms after the verb. (We cannot draw conclusions about the difference between the two analyses, however.) Around the onset of the connective, the trials in which the NP<sub>1</sub> was fixated more often had a causal NP<sub>1</sub> argument than the trials in which the NP<sub>1</sub> was not fixated. To illustrate: 600 ms after the onset of the connective, 85.3% of the trials in which the NP<sub>1</sub> was fixated were trials with a causal NP<sub>1</sub> argument, compared to 77.4% of the trials in which the NP<sub>1</sub> was not fixated.

These exploratory analyses suggest that when the NP<sub>1</sub> is likely to be rementioned in an explanation, it may already be more salient soon after the onset of the verb, compared to when it is unlikely to be rementioned. The effect is small, but striking, particularly given the VSO order used in our experiments: Briefly after the onset of the verb, the structure of the main clause is still unclear, much less the kind of clause that might follow. We found no such early effect for the NP<sub>2</sub>, but the difference between the effect for the NP<sub>1</sub> and NP<sub>2</sub> was small (and the two measures are not independent). The eye-movement record also predicts, starting around the onset of the connective, whether the NP<sub>1</sub> argument of the verb is causal or not. This supports the conclusion from the main analyses that listeners are indeed sensitive to the semantic structure of the verb at an early point during the processing of the *because* clause. The eye-movements seem to predict rementions less well than the causality of the thematic role of the NP<sub>1</sub> at this point, but the current analysis cannot compare the two. We leave it to future research to test the outcomes of these exploratory analyses in a confirmatory way.

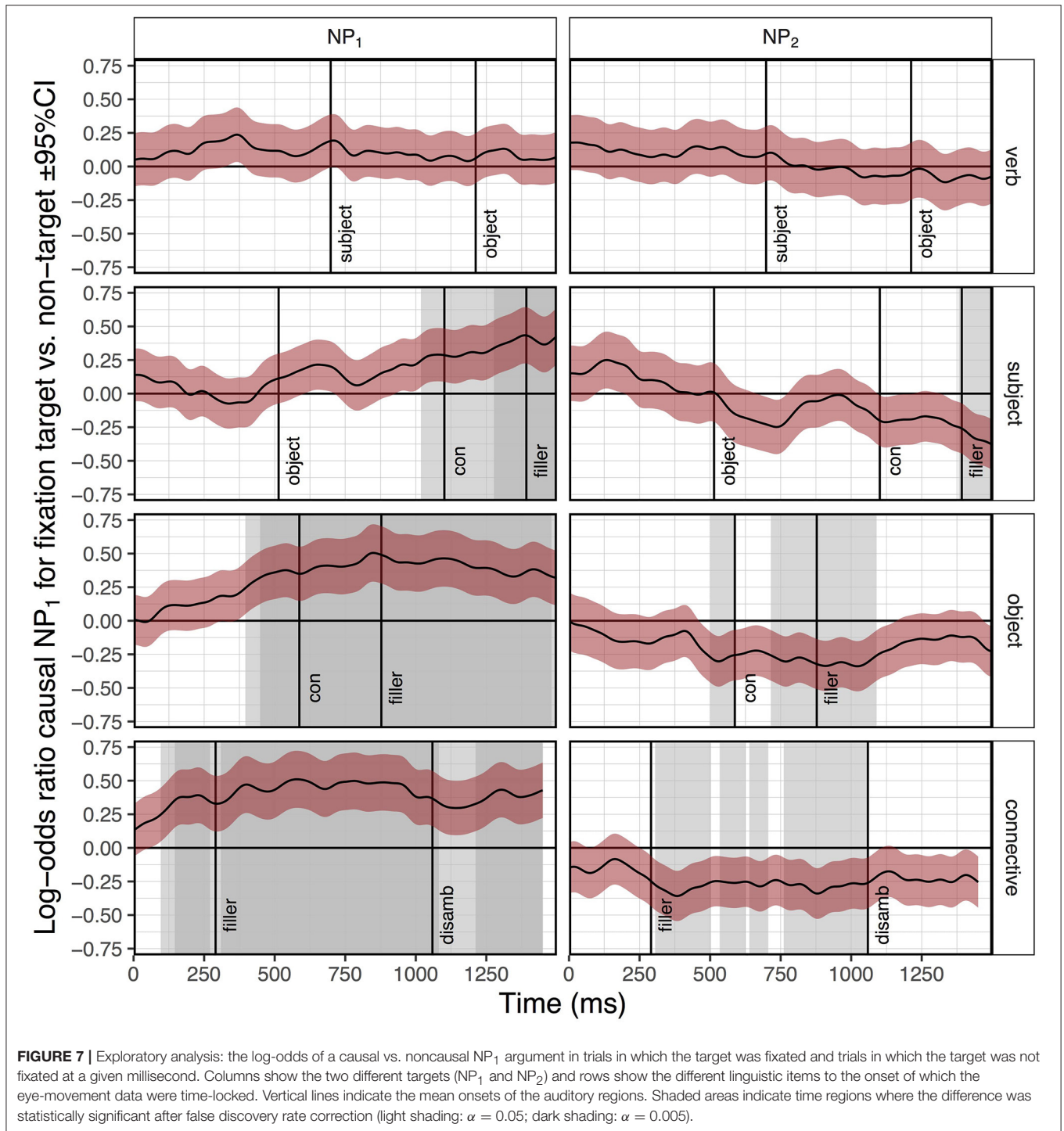


## GENERAL DISCUSSION

We performed two visual world experiments in which IC verbs were embedded in a discourse context that influenced remention biases. IC bias was a significant predictor of looks toward the NP<sub>1</sub> character among items with an NP<sub>2</sub> biased Discourse Context in Experiment 1. When the current results are considered in

combination with results from previous visual world studies on IC, we can conclude that NP<sub>1</sub> biased verbs can lead to a stronger tendency in listeners to establish a coreference relation between the pronoun and the NP<sub>1</sub>, relative to NP<sub>2</sub> verbs, but the effect is not very strong—it only held in one Discourse Context condition in Experiment 1, and it did not show in Experiment 2. There was no main effect of discourse context, so we cannot





**FIGURE 7 |** Exploratory analysis: the log-odds of a causal vs. noncausal NP<sub>1</sub> argument in trials in which the target was fixated and trials in which the target was not fixated at a given millisecond. Columns show the two different targets (NP<sub>1</sub> and NP<sub>2</sub>) and rows show the different linguistic items to the onset of which the eye-movement data were time-locked. Vertical lines indicate the mean onsets of the auditory regions. Shaded areas indicate time regions where the difference was statistically significant after false discovery rate correction (light shading:  $\alpha = 0.05$ ; dark shading:  $\alpha = 0.005$ ).

reject the null hypothesis that listeners initially ignore potentially relevant information from the discourse context during pronoun resolution. Van den Hoven and Ferstl's (2018) story completion study has shown that discourse context does influence remention biases. However, there is no evidence that the same information that is used to guide the search for an appropriate topic to continue the story is immediately used to interpret the pronoun.

Recall that even in the story continuation study, only a subset of stories were numerically NP<sub>1</sub> biased in the NP<sub>1</sub> biased story version and numerically NP<sub>2</sub> biased in the NP<sub>2</sub> biased story version, although the stories were selected on the basis of the difference in remention bias between the two conditions. A larger selection of story pairs of which both versions elicit a bias in the predicted direction might lead to an effect of discourse

context, although a cursory glance at our data does not reveal a much stronger discourse context effect for these stories. Further research using a larger selection of story pairs of which both elicit a bias in the predicted direction would be informative.

It might be argued that the fact that IC bias was involved in an interaction with discourse context is evidence in favor of the world knowledge account: If discourse context had no bearing on pronoun resolution immediately after the onset of the connective, only a main effect of IC bias should be observed, and no interaction. And it is conceivable, under a world knowledge account, that different combinations of verbs and discourse manipulations lead to different effects. However, the world knowledge account did not predict an interaction, and in fact it is difficult to conceive of an account that would predict different interaction effects for the two experiments, given the minimal difference between them. The interaction between IC Bias: 1 and Discourse Context: NP<sub>1</sub> found in Experiment 1 was marginally significant in Experiment 2, but **Figures 2, 4** show that in Experiment 1 it was driven mainly by the NP<sub>1</sub> verbs, whereas in Experiment 2 it was mainly driven by the NP<sub>2</sub> verbs. All of our conclusions are based on the auxiliary hypothesis that listeners tend to look at the character that they believe is being (or going to be) mentioned. Given this assumption, it is not clear why participants should look at the NP<sub>2</sub> character more than at the NP<sub>1</sub> character during NP<sub>1</sub> biased stories with NP<sub>1</sub> verbs (Experiment 1) or equibaised verbs (Experiment 2).

Below, we discuss two possible causes of the observed interaction effects: (i) there are neither true main effects of discourse context nor true interaction effects; all observed effects involving discourse context are simply due to measurement error; and (ii) there is a true main effect of discourse context and no interaction, but a more complex auxiliary hypothesis is needed to account for the observed effects.

We cannot rule out the possibility that the interaction effects are simply due to noise in the data, and discourse context does not actually rapidly influence pronoun resolution. The target sentence was taken from the same audio file across discourse conditions, so any differences between the conditions are either due to the influence of the discourse context or to sampling error or measurement error. As **Figures 2, 4** show, in most cases looking patterns did not differ between discourse conditions a few hundred milliseconds before the average onset of the verb, but for NP<sub>2</sub> verbs in Experiment 1 and for equibaised verbs in Experiment 2, there seems to be a bias in favor of the NP<sub>1</sub> during NP<sub>1</sub> biased stories compared to NP<sub>2</sub> biased stories, starting before the onset of the verb. Such a prior bias could be due to a chance bias in the sampling of combinations of stories and listeners.

The second possibility is that the auxiliary hypothesis is false (or at least too simple). Which part of a scene people fixate is determined by more than just the linguistic input (e.g., Tatler et al., 2011). One well-known effect in research on visual attention is the “inhibition of return” (IOR) effect (Posner and Cohen, 1984). After attention has been oriented to a target area in the scene and subsequently moved away from the target area, response times to stimuli in the target area increase. This effect lasts for several seconds (Klein, 2000). Turning to our data,

in some cases (NP<sub>1</sub> verbs in Experiment 1; NP<sub>1</sub> verbs and equibaised verbs in Experiment 2) there are more looks toward the NP<sub>1</sub> in the NP<sub>1</sub> discourse condition compared to the NP<sub>2</sub> discourse condition shortly after the onset of the verb, but later in the sentence there is a relative decrease in looks toward the NP<sub>1</sub> (Note, however, that there was no relative decrease following the initial increase in NP<sub>1</sub> looks for items with NP<sub>2</sub> verbs in Experiment 1.). This observation is in line with the IOR effect. It is possible that the discourse context already influences eye-movements rapidly after the onset of the verb, and other factors, including IOR, differentially affect the looking patterns in each discourse condition after that time. Unfortunately, we cannot disentangle the possible effects of IOR and discourse context with our current design.

Although there was little doubt before our experiments that IC has an early effect on pronoun resolution in a single sentence context (or a context of two sentences; Pyykkönen and Järvikivi, 2010), our experiments show that, at least in some cases, the effect of IC also holds up when the sentence is embedded in a larger discourse context. This corroborates the method of using longer discourse contexts in pronoun resolution visual world studies: compared to single sentence or two sentence studies there might be more noise in the eye-movements due to IOR and participants making different inferences, but not so much noise that it is impossible to find effects that have been well established in previous research. Our second exploratory analysis suggested an attentional bias toward more causally potent NP<sub>1</sub> arguments compared to less causally potent NP<sub>1</sub> arguments from the beginning of the *because* clause. This suggests a mechanism for the effect of IC bias: Listeners may use this “explicit” causal information (Hartshorne et al., 2015) as an initial heuristic in pronoun resolution.

Apart from the fact that there was a larger discourse context in our experiments, there was another difference between our studies and the earlier visual world studies on IC concerning word order. In all studies except ours, the verb came after its subject or object argument (OVS order was only used in Järvikivi et al., 2017, Experiment 1). We used VSO word order because a lack of any coherence marking indicating the relative time of the target eventuality with respect to the preceding events would be infelicitous in many of the stories (e.g., **Table 1**). And after this sentence-initial coherence marking the word order can only be VSO because German is a V2 language. Assuming that the sentence is processed incrementally, there should be no difference between sentences with VSO and SVO order in terms of the information available to the listener after the second element has been processed, i.e., before the onset of the object. However, the fact that our exploratory analysis suggested a difference in remention bias depending on where listeners looked early after the onset of the verb is striking, since the argument structure of the verb was at that point unclear, although it is not impossible that listeners were able to predict the referents of the verb’s subject and object arguments before they were presented.

Despite the fact that stories were selected on the basis of the effect of discourse context bias on rementions, discourse context only affected looks sporadically, in a different subset of

stories for each experiment, and in unpredicted ways. We took a “shotgun” approach to investigating the effect of discourse context, employing different kinds of manipulations in the same study, and sometimes multiple manipulations within one story pair. As a reviewer suggests, a possible reason for the lack of a clear effect of discourse context is that different discourse manipulations affect online pronoun resolution in different ways, despite affecting rementions in the same way. A *post-hoc* exploration of the effects of the different types of discourse manipulations (sincerity, well-informedness, covariation, and intendedness) showed that the only discourse manipulation that reliably affected looks to the NP<sub>1</sub> character in the predicted direction across the two experiments was sincerity (the type of manipulation employed in the story pair in **Table 1**), and this was only the case for NP<sub>2</sub> verbs. Since there were 10 combinations of IC bias categories and discourse manipulations, the finding that one of them seemed to show a reliable effect should be interpreted with caution. That said, manipulating sincerity information may be a good starting point for future research investigating the effects of different types of discourse manipulations in more detail.

If the auxiliary hypothesis is correct, the lack of an effect of discourse context suggests that the processes involved in establishing a coreference relation between the ambiguous pronoun and one of the antecedents are quite different from the processes involved in choosing an appropriate topic to continue a story: Discourse context influences the latter much more strongly than the former, at least for some types of discourse manipulations. There might be discourse manipulations (perhaps different from the ones we employed) that do reliably affect eye-movements during pronoun resolution. This would be in line with other studies that have shown the effect of discourse context on processing (Nieuwland and Van Berkum, 2006; Rapp and Gerrig, 2006; Kehler and Rohde, 2018). However, our two experiments have shown that, in general, a change in remention bias caused by these discourse manipulations does not necessarily translate into a change in eye-movements during processing, whereas changes in remention bias caused by the IC verb did predict changes in eye-movements in at least one of our experiments.

In conclusion, it has been shown in earlier work that discourse context can alter the remention biases associated with IC verbs

in story completion (Van den Hoven and Ferstl, 2018), but we have found little evidence that these same manipulations of the discourse context also rapidly affect pronoun resolution in the same way. We did find some evidence of the use of IC bias during pronoun resolution in a larger discourse context, giving support to the lexical semantic account. Future research is needed to address two possibilities raised by our exploratory analyses. One is that expectations of rementions affect looks at a very early point, and, in combination with regularities in scene viewing behavior, lead to differential looking patterns after that point. The other is that pronoun resolution is initially guided solely by the causal potency of the NP<sub>1</sub> argument.

## AUTHOR CONTRIBUTIONS

EvdH contributed design of the study, performed the experiments and the statistical analyses, and wrote the first draft of the manuscript. EvdH and EF contributed conception of study, contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

This project was funded by the German Research Foundation (DFG) RTG 1624 Frequency Effects in Language. The paper processing charge was funded by the DFG and the Albert Ludwigs University Freiburg through the funding program Open Access Publishing.

## ACKNOWLEDGMENTS

We thank Lukas Diestel for assistance in the creation and recording of the stimuli, and Anne Cutler, Florian Hintz, Anne Mickan, Lars Konieczny and Arnout Koornneef for helpful discussion at various stages of the project. We thank reviewers Joshua Hartshorne and Alan Garnham for their helpful comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2018.00053/full#supplementary-material>

## REFERENCES

- Altmann, G. T., and Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247–264. doi: 10.1016/S0010-0277(99)00059-1
- Au, T. K. F. (1986). A verb is worth a thousand words: the causes and consequences of interpersonal events implicit in language. *J. Mem. Lang.* 25, 104–122. doi: 10.1016/0749-596X(86)90024-0
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, R. H. (2015b). *Parsimonious Mixed Models*. Available online at: <http://arxiv.org/abs/1506.04967>
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015a). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B.* 57, 289–300.
- Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* 29, 1165–1188. doi: 10.1214/aos/1013699998
- Bott, O., and Solstad, T. (2014). “From verbs to discourse: a novel account of implicit causality,” in *Meaning and Understanding Across Languages: Studies in Theoretical Psycholinguistics*, eds B. Hemforth, B. Mertins, and C. Fabricius-Hansen (Cham: Springer), 213–251.
- Corrigan, R. (2001). Implicit causality in language: Event participants and their interactions. *J. Lang. Soc. Psychol.* 20, 285–320. doi: 10.1177/0261927X01020003002

- Corrigan, R., and Stevenson, C. (1994). Children's causal attributions to states and events described by different classes of verbs. *Cogn. Dev.* 9, 235–256.
- Cozijn, R., Commandeur, E., Vonk, W., and Noordman, L. G. M. (2011a). The time course of the use of implicit causality information in the processing of pronouns: a visual world paradigm study. *J. Mem. Lang.* 64, 381–403. doi: 10.1016/j.jml.2011.01.001
- Cozijn, R., Noordman, L. G. M., and Vonk, W. (2011b). Propositional integration and world-knowledge inference: processes in understanding because sentences. *Discourse Process.* 48, 475–500. doi: 10.1080/0163853X.2011.594421
- Crawley, R. A., Stevenson, R. J., and Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *J. Psycholinguist. Res.* 19, 245–264. doi: 10.1007/BF01077259
- Crinean, M., and Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Lang. Cogn. Process.* 21, 636–648. doi: 10.1080/01690960500199763
- Croft, W. (2012). *Verbs: Aspect and Causal Structure*. Oxford: OUP.
- Featherstone, C. R., and Sturt, P. (2010). Because there was a cause for concern: an investigation into a word-specific prediction account of the implicit-causality effect. *Q. J. Exp. Psychol.* 63, 3–15. doi: 10.1080/17470210903134344
- Ferstl, E. C., Garnham, A., and Manouilidou, C. (2011). Implicit causality bias in English: a corpus of 300 verbs. *Behav. Res. Methods* 43, 124–135. doi: 10.3758/s13428-010-0023-2
- Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. London: Sage.
- Frederiksen, J. R. (1981). Understanding anaphora: rules used by readers in assigning pronominal referents. *Discourse Process.* 4, 323–347. doi: 10.1080/01638538109544525
- Garnham, A., Traxler, M., Oakhill, J., and Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *J. Mem. Lang.* 35, 517–543. doi: 10.1006/jmla.1996.0028
- Garvey, C., and Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry* 5, 459–464.
- Gernsbacher, M. A., and Hargreaves, D. J. (1988). Accessing sentence participants: the advantage of first mention. *J. Mem. Lang.* 27, 699–717. doi: 10.1016/0749-596X(88)90016-2
- Greene, S. B., and McKoon, G. (1995). Telling something we can't know: experimental approaches to verbs exhibiting implicit causality. *Psychol. Sci.* 6, 262–270. doi: 10.1111/j.1467-9280.1995.tb00509.x
- Guerry, M., Gimenes, M., Caplan, D., and Rigalleau, F. (2006). How long does it take to find a cause? an on-line investigation of implicit causality in sentence production. *Q. J. Exp. Psychol.* 59, 1535–1555. doi: 10.1080/17470210500269105
- Hartshorne, J. K. (2014). What is implicit causality? *Lang. Cogn. Neurosci.* 29, 804–824. doi: 10.1080/01690965.2013.796396
- Hartshorne, J. K., O'Donnell, T. J., and Tenenbaum, J. B. (2015). The causes and consequences explicit in verbs. *Lang. Cogn. Neurosci.* 30, 716–734. doi: 10.1080/23273798.2015.1008524
- Hartshorne, J. K., and Snedeker, J. (2013). Verb argument structure predicts implicit causality: the advantages of finer-grained semantics. *Lang. Cogn. Process.* 28, 1474–1508. doi: 10.1080/01690965.2012.689305
- Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geysen, A., et al. (2011). dlexDB - eine lexikalische Datenbank für die psychologische und linguistische Forschung [dlexDB - a lexical database for psychological and linguistic research]. *Psychol. Rundsc.* 62, 10–20. doi: 10.1026/0033-3042/a000029
- Hobbs, J. R. (1979). Coherence and coreference. *Cogn. Sci.* 3, 67–90. doi: 10.1207/s15516709cog0301\_4
- Itzhak, I., and Baum, S. R. (2015). Misleading bias-driven expectations in referential processing and the facilitative role of contrastive accent. *J. Psycholinguist. Res.* 44, 623–650. doi: 10.1007/s10936-014-9306-6
- Järvikivi, J., van Gompel, R. P., and Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *J. Psycholinguist. Res.* 46, 525–550. doi: 10.1007/s10936-016-9451-1
- Kamide, Y., Altmann, G. T., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8
- Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI.
- Kehler, A., Kertz, L., Rohde, H., and Elman, J. L. (2008). Coherence and coreference revisited. *J. Semant.* 25, 1–44. doi: 10.1093/jos/ffm018
- Kehler, A., and Rohde, H. (2018). Prominence and coherence in a Bayesian theory of pronoun interpretation. *J. Pragm.* doi: 10.1016/j.pragma.2018.04.006. [Epub ahead of print].
- Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symp. Motiv.* 15, 192–238.
- Kelley, H. H. (1973). The processes of causal attribution. *Am. Psychol.* 28, 107–128. doi: 10.1037/h0034225
- Klein, R. M. (2000). Inhibition of return. *Trends Cogn. Sci.* 4, 138–147. doi: 10.1016/S1364-6613(00)01452-2
- Koornneef, A., Dotlacil, J., van den Broek, P., and Sanders, T. (2015). The influence of linguistic and cognitive factors on the time course of verb-based implicit causality. *Q. J. Exp. Psychol.* 69, 455–481. doi: 10.1080/17470218.2015.1055282
- Koornneef, A. W., and Sanders, T. J. (2013). Establishing coherence relations in discourse: the influence of implicit causality and connectives on pronoun resolution. *Lang. Cogn. Process.* 28, 1169–1206. doi: 10.1080/01690965.2012.699076
- Koornneef, A. W., and Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *J. Mem. Lang.* 54, 445–465. doi: 10.1016/j.jml.2005.12.003
- Lafrance, M., Brownell, H., and Hahn, E. (1997). Interpersonal verbs, gender, and implicit causality. *Soc. Psychol. Q.* 60, 138–152. doi: 10.2307/2787101
- Long, D. L., and De Ley, L. (2000). Implicit causality and discourse focus: the interaction of text and reader characteristics in pronoun resolution. *J. Mem. Lang.* 42, 545–570. doi: 10.1006/jmla.1999.2695
- Majid, A., Sanford, A. J., and Pickering, M. J. (2006). Covariation and quantifier polarity: what determines causal attribution in vignettes? *Cognition* 99, 5–51. doi: 10.1016/j.cognition.2004.12.004
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024
- McKoon, G., Greene, S. B., and Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *J. Exp. Psychol.* 19, 1040–1052. doi: 10.1037/0278-7393.19.5.1040
- Nieuwland, M. S., and Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *J. Cogn. Neurosci.* 18, 1098–1111. doi: 10.1162/jocn.2006.18.7.1098
- Pickering, M. J., and Majid, A. (2007). What are implicit causality and consequentiality? *Lang. Cogn. Process.* 22, 780–788. doi: 10.1080/01690960601119876
- Posner, M. I., and Cohen, Y. (1984). “Components of visual orienting,” in *Attention and Performance, Vol. X*, eds, H. Bouma and D. Bouwhuis. (Hilldale, NJ: Erlbaum), 531–556.
- Pykkönen, P., and Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Exp. Psychol.* 57, 5–16. doi: 10.1027/1618-3169/a000002
- Rapp, D. N., and Gerrig, R. J. (2006). Predispositions for narrative outcomes: the impact of story contexts and reader preferences. *J. Mem. Lang.* 54, 54–67. doi: 10.1016/j.jml.2005.04.003
- Roberts, C. (2012). Information structure in discourse: towards an integrated formal theory of pragmatics. *Seman. Pragm.* 5, 1–69. doi: 10.3765/sp.5.6
- Scheepers, C., Hemforth, B., and Konieczny, L. (2000). “Linking syntactic functions with thematic roles: psych-verbs and the resolution of subject-object ambiguity,” in *German Sentence Processing*, eds, B. Hemforth and L. Konieczny (Dordrecht: Kluwer), 95–135.
- Schlechter, J., and Felser, C. (2016). *Can Non-native Speakers Use Implicit Causality Information as a Predictive Cue?* Poster presented at the Conference on Multilingualism, Ghent.
- Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *J. Verbal Learning Verbal Behav.* 13, 272–281. doi: 10.1016/S0022-5371(74)80064-2
- Smyth, R. (1994). Grammatical determinants of ambiguous pronoun resolution. *J. Psycholinguist. Res.* 23, 197–229. doi: 10.1007/BF02139085
- Stewart, A. J., Pickering, M. J., and Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *J. Mem. Lang.* 42, 423–443. doi: 10.1006/jmla.1999.2691



- Tatler, B. W., Hayhoe, M. M., Land, M. F., and Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting saliency. *J. Vis.* 11, 1–23. doi: 10.1167/11.5.5
- Van Berkum, J. J., Koornneef, A. W., Otten, M., and Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171. doi: 10.1016/j.brainres.2006.06.091
- Van den Hoven, E., and Ferstl, E. C. (2017). Association with explanation-conveying constructions predicts verbs' implicit causality biases. *Int. J. Corp. Ling.* 22, 521–550. doi: 10.1075/ijcl.16121.hov
- Van den Hoven, E., and Ferstl, E. C. (2018). Discourse context modulates the effect of implicit causality on rementions. *Lang. Cogn.* 1–34. doi: 10.1017/langcog.2018.17
- Wittenberg, E., Khan, M., and Snedeker, J. (2017). Investigating thematic roles through implicit learning: evidence from light verb constructions. *Front. Psychol.* 8:1089. doi: 10.3389/fpsyg.2017.01089

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

Copyright © 2018 van den Hoven and Ferstl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.