Comprehending Texts and Pictures: Interactions Between Linguistic and Visual Processes in Children and Adults

Jan Engelen

ISBN: 978-90-5335-888-7

Cover drawings by E.C. Eielts
Cover design by J.A.A. Engelen
Lay-out by J.A.A. Engelen
Printed by Ridderprint B.V., Ridderkerk, the Netherlands

Comprehending Texts and Pictures:
Interactions between linguistic and visual processes in children and adults

Het begrijpen van teksten en afbeeldingen:
Interacties tussen talige en visuele processen in kinderen en volwassenen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 11 september 2014 om 11.30 uur

door

Johannes Antonius Adrianus Engelen
geboren te Terheijden

Promotiecommissie

Promotor:

Prof.dr. R.A. Zwaan

Overige leden:

Prof.dr. J.P. Magliano

Dr. K. Dijkstra

Dr. D. Pecher

Copromotoren:

Dr. S. Bouwmeester

Dr. A.B.H. de Bruin

Contents

Chapter 1

Introduction

On Monday morning, the birthday boy was walking to school with another boy. They were passing a bag of potato chips back and forth and the birthday boy was trying to find out what his friend intended to give him for his birthday that afternoon. Without looking, the birthday boy stepped off the curb at an intersection and was immediately knocked down by a car. He fell on his side with his head in the gutter and his legs out in the road. His eyes were closed, but his legs moved back and forth as if he were trying to climb over something. His friend dropped the potato chips and started to cry. The car had gone a hundred feet or so and stopped in the middle of the road. The man in the driver's seat looked back over his shoulder. He waited until the boy got unsteadily to his feet. The boy wobbled a little. He looked dazed, but okay. The driver put the car into gear and drove away.

These are the events that take place at the beginning of 'A Small, Good Thing', a short story by Raymond Carver (1983). Later, the boy, whose name is Scotty, falls unconscious and is hospitalized. During the hours she visits home to rest, his mother receives a series of strange calls that turn out to be the baker reminding her about the birthday cake she had ordered. Despite these bleak circumstances, it is a work of fiction I thoroughly enjoyed. However, the reason I cite this mildly upsetting passage is that it serves to illuminate some themes that are important throughout the present thesis, in which I investigate text comprehension in children and adults. What mental processes took place while reading that passage, and how did these lead to understanding?

Broadly speaking, the reader faces the challenge of constructing meaning at two hierarchically related levels: at the micro-level, the meanings of the words need to be combined, under the guidance of syntax, into a mental representation of the state of affairs described in a sentence. At the macro-level, these mental representations need to be integrated to form a coherent structure. Take the sentence *The man in the driver's seat looked back over his shoulder.* Among other things, the reader needs to find appropriate referents for the nouns *man*, *driver's seat*, and *shoulder*, as well as for the verb *looked back*. The reader also needs to understand that *his shoulder* is coreferential with *the man*, that *the man* is the agent performing the act of moving his head, and that this event took place in the past. To build coherence across sentences, the reader needs to infer that *the man in the driver's seat* points back to *the car* in the previous sentence, and that the act of looking is triggered by the accident just moments ago.

Within this framework, the research in this thesis focuses on three sets of questions. First, how do the comprehension processes at the macro-level constrain those at the micro-level? Meaning construction at the micro-level has frequently been studied in single sentences, but rarely embedded within a discourse context – a potentially dangerous scenario (Clark, 1997). The processes the reader engages in to comprehend *The man in the driver's seat looked back over his shoulder* might be different in the context of the story than in isolation.

Second, how do children construct meaning at the micro- and the macro-level? While several of these processes are extensively documented for adults, they are not for children. Thus, there is a descriptive gap. But more importantly, developmental findings have ramifications for theory. A good theory of reading comprehension is plausible in light of developmental and individual differences. A comprehensive theory – which is far beyond the scope of this thesis – also predicts them.

Third, what is the role of the visual context on comprehension? Just as a discourse context imposes constraints on how the meaning of a sentence is represented, so might information from other modalities. In particular, I focus on the role of pictorial illustrations. Do specific visual details transfer into readers' mental representations, and can we use looks toward pictures to uncover some of the processes of discourse comprehension? The next sections will present a theoretical outlook on comprehension and motivate these research questions in more detail.

Comprehending Connected Discourse: The Event-Indexing Model

A generally accepted view is that readers construct situation models, that is, mental representations of the events that are described in the text, rather than of the text itself (Bower & Morrow, 1992; Bransford, Barclay, & Franks, 1972; Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Situation model construction proceeds in an incremental fashion: each incoming unit of information is immediately interpreted and integrated with the existing model. Zwaan and Radvansky (1998) refer to this process as *updating*. In doing so, the reader actively seeks to make the new information cohere with previously comprehended information. According to the event-indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995), at least five dimensions of coherence are monitored: time, space, protagonist, intentionality, and causation. Let us take a closer look at how this might have happened for the above passage.

The first sentence places the situation on a Monday morning. Because there are no explicitly signaled temporal shifts in the passage (e.g., *an hour later* or *on Tuesday*), the reader processes the events under the default assumption that they are all temporally contiguous (Zwaan, 1996). As a consequence, all entities in the discourse model are kept accessible in working memory, or *foregrounded*. Only a few units of information can be in the focus of attention at a given time. Therefore, the reader has to rely on a long-term memory structure from which information can be efficiently retrieved. The presence and absence of such temporal markers can be seen as *heuristics* as to what information should be kept accessible and what information may be deactivated. After all, the further away from the narrative *now*, the less likely it is that particular information will continue to be relevant.

The first sentence also introduces a spatial setting: the way to school, which is later narrowed down to a particular intersection. As with time, information related to the same spatial scene – unlike information from different locations – remains accessible in the discourse model (Glenberg, Meyer, & Lindem, 1987). The situation is described from the perspective of an external observer, possibly causing the reader to adopt that view on the events, as if she were witnessing them herself (Franklin & Tversky, 1990). On this more fine-grained level, the 'mental camera' moves a couple of times, first from the birthday boy to the other boy, then to the car and back to the birthday boy, and then once more to the car. As it does so, the reader's attention is drawn to novel elements in the scene, such as the birthday boy's condition and the other boy's and the driver's reaction.

Another dimension of coherence is that of protagonist. Simply put, narrative discourse is about characters and their actions (Sanford & Garrod, 1981), so naturally, readers are engaged in tracking what each new sentence predicates of whom. In our example, the first sentence introduces two story characters. One is immediately foregrounded in terms of importance by the definite article *the* and the relatively elaborate description *birthday boy*. Also, the order in which the two characters are mentioned gives the reader a clue about their relative importance: first-mentioned entities tend to be the subjects in the subsequent discourse (Gernsbacher & Hargreaves, 1988). Thus, the birthday boy is established as the protagonist and the most likely referent of anaphoric expressions such as *he* and *his* that occur in later sentences (Morrow, 1985).

Closely related to protagonist, intentionality is another dimension of coherence – one might say that the goals and plans of characters are the catalysts of narratives (Schank & Abelson, 1977). The second sentence contains an explicit goal statement: the birthday boy wants to find out what he will get for a present. This makes it likely that the narrative will go on to describe what the boy does to achieve that goal. The reader remembers that goal until it is completed (or, as in this case, gets tragically interrupted) (Lutz & Radvansky, 1997). Goals can also be inferred. For instance, when the car stops, the reader is likely to deduce that the driver wants check on the boy (but learns that the driver decides otherwise).

Finally, the reader is concerned with explaining the events in text. To do so, causal inferences are made between the information in the current sentence and information stated earlier in the text or residing in the reader's prior knowledge (Trabasso & van den Broek, 1985). For instance, the boy falls to his side with his head in the gutter because of the impact of the car. The other boy's reaction of crying and dropping the chips is caused by his shock over the accident. Note that the passage contains no explicit causal markers. It would have been fine to restate the final sentence as '*Therefore*, the driver put the car into gear and drove away', but this causal link is so routinely inferred that the signaling device is unnecessary.

These interactions between the text and the reader lead to a memory structure in which the individual events are firmly interconnected – especially if they cohere on one or more situational dimensions (e.g., because they take place within the same temporal interval, or are part of the same causal chain). This episodic memory structure, which Zwaan and Radvansky (1998) call the *complete model*, can be considered the end product of narrative comprehension.[1]

Comprehending Individual Events: Perceptual Symbol Systems

So far, I have presented a common, fairly uncontroversial analysis of the coherence relations that readers routinely monitor in a given passage of text. However, in the late 1990s researchers became increasingly aware that one aspect had not been addressed satisfactorily: how do readers understand the events themselves? Until then, mainstream theories of meaning making (e.g., Kintsch, 1998; Newell & Simon, 1972) posited that readers store the events described in a text as propositions. By definition, these are abstract, stripped of modality-specific information, and arbitrarily related to their external referents. Consider the final event in the passage: the car driving away. This particular node in the situation model might be represented propositionally as DRIVE[CAR, AWAY]. The arguments in this proposition, CAR and AWAY, could in turn be represented as a list of features, such as [VEHICLE, MOTORIZED]. A legitimate follow-up question would be how the meaning of the features in this list is represented. This would just be another list of features – from which point the question could be repeated *ad infinitum*. This conundrum is known under various names, such as the 'Chinese room problem' (Searle, 1980), the 'symbol merry-go-round', and the 'grounding problem' (Harnad, 1990). The essence of the problem is this: for linguistic symbols to have any meaning at all, they must at some point map onto something outside the system.

An elegant solution for this problem is provided by Barsalou's (1999) perceptual symbol systems, which propose that the meaning of linguistic structures is grounded in a multimodal record of neural activation. So, the instantiation of the meaning of *car*, in the present narrative context, lies in the reactivation of relevant visual and, to a lesser extent, auditory, olfactory, haptic, and motor memory traces. Importantly, perceptual symbols are still symbols – things standing in for other things, that can be combined and manipulated to form higher-order representations (Deacon, 1997) – but they do keep an analog relationship to the sensorimotor experience that gave rise to them. Following this proposal, a host of empirical research has shown that language comprehension involves the activation of modality-specific systems. For instance, pictures can be named faster if

[1] There is good reason to disagree that such a memory representation is also the *goal* of comprehension. We do not read stories, listen to news reports, and study recipes just to remember them. Rather, comprehension is for action (Glenberg, 1997). However, this does not discredit the existence of long-term memory representations, which may be highly functional in guiding behavior (see Mar & Oatley, 2008).

their shape matches that implied by a sentence read previously (Zwaan, Stanfield, & Yaxley, 2002), sensibility judgments are given faster if the direction of movement implied in the sentence matches the direction of movement required to execute the response (Glenberg & Kaschak, 2002), and aurally presented words from a written dialog are recognized faster when the voice matches the gender of the character (Kurby, Magliano, & Rapp, 2009). Apart from offering a solution to the grounding problem, perceptual symbol systems are appealing in that they seem well-suited to explain the feeling of being part of the imaginary situation that readers often experience with well-written text – also known as *transportation* (Gerrig, 1993; Green & Brock, 2000). The studies described in this thesis are mostly framed within this *embodied* view on language comprehension, and on cognition more generally.

Assumptions Regarding Comprehension: Lower-Order Skills

In the analysis of the passage from 'A Small, Good Thing', several things were taken for granted. First, it was assumed that the reader knew all the words in the passage – that is, each lexical item, such as *intersection* and *dazed*, could directly be connected to a referent in long-term memory, or, more appropriately, grounded in a multimodal record of neural activation. Vocabulary, both in terms of breadth (how many words are known – typically measured by how many pictures of a set can be accurately named) and depth (how well the meanings are known – typically measured by how many defining features of a given word can be listed within a limited amount of time) is an important predictor of reading outcomes (Ouellette, 2006). An estimate is that the reader needs to know at least 98% of the word tokens in a text to achieve a reasonable level of comprehension (Hu & Nation, 2000).

Second, it was assumed that reading of the passage went fluently, and that the reader could focus his or her attention on constructing meaning, rather than identifying individual words. According to the *simple view of reading* (Gough & Tunmer, 1986), what the reader takes away from a text is predicted by the product of text decoding skills and general comprehension skills. This means that when either is zero, there is essentially no comprehension. Therefore, when studying specific comprehension abilities, it can be practical to rule out influences of decoding skill and present a text in spoken form, although even then individual differences in word recognition ability may factor into the outcome. According to the *lexical quality hypothesis* (Perfetti & Hart, 2002), the efficiency of decoding is closely related to lexical knowledge. A representation is said to be of high quality when the orthographic (i.e., spelling), phonological (i.e., pronunciation), and semantic (i.e., meaning) representation of a word are tightly connected. As a result, retrieval of one type of information leads to the rapid and synchronous activation of the other types of information. For example, the word *intersection*, when read by an individual in possession of a high-quality representation,

will activate the correct pronunciation, as well as the appropriate meaning, in a unitary experience. However, if the representation is of poor quality, reading the word will not consistently activate the correct pronunciation, and trigger closely related orthographic and semantic forms, causing confusion with words like *interjection* or *inner section.* As a result, comprehension will not be optimal. This account, more than accounts that cast poor decoding skills as a simple bottleneck of processing capacity (e.g., Perfetti, 1985), highlights the intricate relationship between word reading and comprehension and the extent to which they are intertwined with knowledge.

Third, it was assumed that the reader had a basic understanding of how events relate to each other in the world. For example, a person who is hit by a car is likely to be hurt badly, and witnessing it can be very shocking to bystanders. Without this knowledge, it might be very difficult to make the inferential connections between the events that are necessary to perceive the text as a coherent whole, rather than a set of unrelated statements. Relevant background knowledge is a strong predictor of comprehension of specific texts, even compensating for small vocabularies and non-fluent word reading (Marr & Gormley, 1982).

While these assumptions are warranted, by and large, for adult readers, they are not for developing readers. Indeed, these component skills (i.e., vocabulary, word reading, and general world knowledge) all independently account for substantial variance in comprehension outcomes (Oakhill, Cain, & Bryant, 2003). As of yet, little is known about whether different levels of mastery of these component skills may lead to *qualitative* differences among children's mental representations. This is an interesting topic, which is touched upon in Chapter 2. Generally, however, the study of higher-order comprehension processes is difficult if the researcher has insufficient control over these low-level variables. While it is possible to match research participants on these variables, it is not always possible to construct research materials that are controlled in terms of difficulty without at least somewhat artificial results (Graesser et al., 1997). This puts the researcher in a tough position: to what extent should experimental control be given up in favor of authenticity?

This dilemma becomes especially poignant when we consider a final assumption: the reader processed the passage with the goal of constructing a globally coherent representation, or engaged in what Bartlett (1932) called an *effort after meaning.* In principle, a given text can be processed at any level of depth. On one end of the continuum, there is very shallow comprehension of the sort where inferences are encoded only when they are required for local coherence and when the information needed to support them is immediately accessible (McKoon & Ratcliff, 1992). According to this view, the reader will make the inference that *The car had gone a hundred feet or so and stopped in the middle of the road* refers to the same car that ran into Scotty, but not that the driver only stopped briefly because he was in a hurry. On the other end, there is

hermeneutic interpretation, where readers pay great effort to uncover the possible meanings conveyed by the text. This might amount to searching for a deeper interpretation of the bag of chips, speculating about the role of the other boy, et cetera.[2] Of course, the reader's comprehension goals can vary across situations and with that the amount of inferencing. Key factors determining experimental outcomes in this regard are the quality and length of the texts and the instructions that the reader has received. First, minimalist processing becomes more elaborate when text length increases (e.g., from four to eight sentences; Love & McKoon, 2011). In isolation, the sequence *The boy wobbled a little. He looked dazed, but okay. The driver put the car into gear and drove away* might be perceived as trivial, not prompting much inferential processing. However, with the preceding context about the accident, the reader is likely to have a stronger commitment to developing an elaborate mental representation. Second, mental representations tend to be more fully specified when participants are required to answer a question or rewrite some sentences in their own words, rather than to just read the text (Foertsch & Gernsbacher, 1994). *Standards of coherence* are a useful construct to capture the set of criteria that readers apply to their processing of a specific text with a specific goal (van den Broek, Risden, & Husebye-Hartmann, 1995). I assume that the reader typically has fairly high standards of coherence: some types of inferences, such as the goals of the characters and the causal antecedents of events (e.g., the other boy started to cry because he was shocked), are likely to be made automatically and online, whereas others, such as the author's intent or an event's causal consequence (e.g., the chips might spill on the curb), are rather made strategically and offline (Graesser, Singer, & Trabasso, 1994). With these assumptions in place, the next section outlines the research questions of each individual chapter and gives a short overview of the methodology.

Overview of Studies in this Thesis

Chapter 2 looks at meaning making at the sentence level. Specifically, we asked whether 7 to 12-year-old children ground the events described in short utterances in their own perceptual experiences. To this end, we used the perceptual mismatch paradigm, which has had a notable career in serving the development of theories of grounded language comprehension. Stanfield and Zwaan (2001), for instance, found that readers were faster to respond to a picture of a vertical pencil after they had read *John put the pencil in the cup* than *John put the pencil in the drawer*, and vice versa for a picture of a horizontal pencil. Zwaan et al. (2002) manipulated shape rather than orientation and found the same pattern of results with sentence pairs like *The ranger saw the eagle in the sky* and *The ranger saw the eagle in the nest* followed by a picture of either a flying or a perched eagle. These studies support the notion that the mental representations of

[2] If this seems far-fetched, please take a look at http://tinyurl.com/smallgood (retrieved February 21, 2014).

described events are analogous to their real-world counterparts. We used a set of sentences in which both orientation and shape were manipulated.

There is good reason to perform this type of experiment with developing comprehenders. Given that much of conceptual knowledge has its basis in spatial representations (Mandler, 2010), it would seem reasonable to hypothesize that young children's language comprehension is also grounded in these representations. However, other evidence suggests that perceptual simulations are only constructed when sentences can be efficiently processed (Madden & Zwaan, 2006) and when the reader has sufficient direct experience with the events that are described (Holt & Beilock, 2006). Precisely these requisite skills are significantly less developed in young children than in adults. Thus, it is far less likely that young children also routinely perform detailed mental simulations. If they do, however, despite these constraints, this might have important implications for our outlook on language comprehension: it makes quite a difference whether mental simulation occurs at the endpoint of development (i.e., in experts) or at the origin (i.e., in beginning readers). Furthermore, we addressed the question of whether reading from print, as opposed to listening, may be a bottleneck in performing simulations, by comparing the performance of fluent word readers with that of struggling word readers.

In **Chapter 3**, we scale our research up to connected discourse. It deals with the question of how 6 to 12-year-old children process anaphoric expressions such as *he* and *they*, which are used frequently throughout discourse, but carry very little meaning in and of themselves – the only semantic information encoded in *he* is 'singular masculine entity', while *they* is even less constraining and merely restricts the domain of reference to 'plural entity'. Instead, they are heavily parasitic on the surrounding discourse. Research has shown that the referent of an anaphoric pronoun is typically the most prominent entity at that particular juncture in the discourse. Prominence, in turn, is determined by such things as order-of-mention, subjecthood, and recency (Arnold et al., 2007). We subscribe to the view that these dependencies are not captured in some linguistic rule that needs to be discovered by the child, but rather in probabilistic patterns of co-occurrence in the discourse (Bates & MacWhinney, 1987). It takes much exposure to build a sufficiently rich discourse record. Therefore, we may expect substantial developmental and individual variation in the application of these cues to anaphor resolution.

As discussed earlier in this introduction, comprehenders need to resolve anaphoric pronouns in order to keep a mental record of who and what the text is about. Therefore, poor anaphoric resolution should also lead to a degraded overall representation of the text – which is what we tried to find out by combining an online measure of text processing and an offline measure of memory. Children listened to a 7-minute story and concurrently viewed pictures of its protagonists, while their eye movements were being recorded. Afterward, they received a 15-item oral comprehension test. We hypothesized

that accuracy on this test would correspond to the gaze behavior in response to hearing anaphoric expressions in the story. In particular, good comprehenders should be more likely to fixate the referents of expressions such as *he* and *him* than poor comprehenders. Chapters 2 and 3 highlight two conceptually separable levels of comprehension: that of individual events (i.e., the micro-level) and that of the coherence relations between them (i.e., the macro-level). To date, there is very little work that investigates the interplay between these levels. **Chapter 4**, therefore, is concerned with the role of grounded event representations during discourse comprehension. On the one hand, the two seem to have a mutually reinforcing relationship. Discourse context facilitates elaborate grounded representations (e.g., Kurby & Zacks, 2013). Similarly, without sufficiently detailed, grounded event representations, building coherence across sentences is difficult (Fincher-Kiefer, 2001). On the other hand, given that both, although presumably constituting highly automatized skills in proficient adults, cost *some* cognitive resources, it may be that a strong focus on one interferes with the execution of the other. In that case, what do comprehenders do?

We tested this scenario by introducing conflict between grounding action events and building coherence: participants listened to discourse-embedded sentences while viewing pictures that matched the protagonist of the story, but not the action of the specific sentence, or vice versa. For instance, participants heard *The baker did not hesitate and dove into the water* while seeing an idly standing baker and a diving athlete. Crucially, only one picture could be attended to at a given time. So, an event-internal focus should lead to looks toward the action picture (the athlete), while coherence-driven processing should lead to looks toward the agent picture (the baker). By continuously tracking participants' gaze as the sentence unfolded, we were able to draw conclusions about the relative dominance and the temporal dynamics of these levels of representation during comprehension.

In both Chapters 3 and 4, looks toward pictures were treated as an index of comprehension. But the precise causal role of eye movements across the visual environment remained unclear: do language users look at pictures because they provide relevant visual information, or do pictures have a different function in supporting the comprehension process? Especially when pictures are simplified line drawings, they may not enhance the mental representations that readers and listeners routinely construct. In this context, **Chapter 5** explores whether specific properties of the visual environment influence mental simulations. Suppose the reader had seen a picture of the two boys with their bag of chips prior to reading, and that they were facing left from the reader's perspective. Seeing this picture might provide a salient and easily accessible record of perceptual activation that can be used to ground the meaning of the text. Does the specific orientation make the reader more likely to represent the accident with the car as taking

place on the left as well? Or are there other principles governing the spatial layout of the imagined situation?

To investigate this, we presented participants with a picture of a character that was facing left or right, after which they read a sentence in which that character was described as moving toward or away from an object, such as *The doctor walked toward the cabinet where he kept the patient's file.* Then they saw a picture of an object, which was located either on the far left or far right of the screen, and were asked to verify if it was mentioned in the sentence. If the prior visual cue and the sentence coalesce in determining where readers direct their visual attention, participants should be faster to verify the object if it was in a location that was congruent with the character's implied moving direction than in a location that was incongruent with it.

The Importance of Studying Text Comprehension

I have saved the discussion of another crucial assumption for last: studying text comprehension is valuable. Why spend years of one's life researching it – or, for that matter, hours reading a lengthy thesis about it? First, understanding language comprises a major part of our daily lives. One might argue that using language to communicate ideas across time and space is one of the most sophisticated skills that humans possess. While language is not necessarily *text*, it rarely comes in the form of isolated words or sentences; rather, it is embedded in a pragmatic and discursive context. As I hope the analysis of just a single paragraph from 'A Small, Good Thing' made clear, successful comprehension requires the coordination of several levels of analysis. As such, text processing offers some unique challenges that cannot be reduced to, say, the domain of memory or linguistics.

Second, a goal of this research is to inform educational practice. Being able to comprehend written and spoken language is an extremely important skill for learning. Whether it comes down to understanding subject matter in history, biology, mathematics, finding relevant information on the internet, or reading and enjoying stories for their own sake: without well-developed comprehension abilities, a child is not fit to meet the demands of an increasingly information-driven society. Not surprisingly, reading skill is a strong predictor of school achievement (e.g., Maughan, 1995; Savolainen, Ahonen, Aro, Tolvanen, & Holopainen, 2008; Spreen, 1987). While many children show an interest in deciphering print, and several prerequisite skills, such as phonological awareness, appear to develop almost spontaneously (Whitehurst & Lonigan, 1998), most children need to be *taught* to read. This holds for lower-order skills (e.g., spelling-to-sound mappings) as much as for higher-order skills (e.g., comprehension strategies). Knowing precisely what skills underlie successful comprehension is a necessary condition for evaluating existing educational approaches and making informed improvements. Existing research-based interventions that target the reader, the text, or the instructional setting, show promise in

changing both the process and the outcome of comprehension (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). Ultimately, their success depends on a correct and full understanding of what constitutes comprehension and the processes that lead up to it.

The focus on the interaction between pictures and text is also highly relevant in this regard. Pictures and text have a long shared history in education (see Smith, 1965). For instance, spelling-to-sound mappings are commonly taught with reference to pictures. Also, storybooks for children typically contain illustrations, and more complex matter is often explained in the context of diagrams, schemas, and figures. To what extent such illustrations impact the mental representations built from text, both at the micro- and the macro-level, is a question that is at the heart of the larger issue of whether they are beneficial to learning.

Chapter 2

Perceptual Simulation in Developing Language Comprehension*

* This chapter has been published as Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, *110*, 659-675.

Abstract

We tested an embodied account of language which proposes that comprehenders create perceptual simulations of the events they hear and read about. In Experiment 1, children (ages 7-13) performed a picture verification task. Each picture was preceded by a pre-recorded spoken sentence describing an entity whose shape or orientation matched or mismatched the depicted object. Responses were faster for matching pictures, suggesting that participants had formed perceptual-like situation models of the sentences. The advantage for matching pictures did not increase with age. Experiment 2 extended these findings to the domain of written language. Participants (ages 7-10) of high and low word reading ability verified pictures after reading sentences aloud. The results suggest that even when reading is effortful, children construct a perceptual simulation of the described events. We propose that perceptual simulation plays a more central in role in developing language comprehension than previously thought.

Consider the sentences *The ranger saw the eagle in the sky* and *The ranger saw the eagle in the nest*. The former refers to an eagle with outstretched wings, while the latter refers to an eagle that has its wings drawn in. Theories of embodied language comprehension predict that when readers process these sentences, their mental representation of the eagle changes accordingly (Zwaan, Stanfield, & Yaxley, 2002). To test this prediction, researchers have used the perceptual mismatch paradigm (Stanfield & Zwaan, 2001), in which a participant reads or listens to a sentence describing a situation, and is subsequently presented with a picture of an object involved in that situation. Critically, the shape or orientation of the depicted object either matches or mismatches the shape or orientation implied by the description. For example, *The ranger saw the eagle in the sky* can be followed by a picture of a flying eagle (match) or a perched eagle (mismatch). It has been found that participants need more time to verify or name mismatching pictures relative to matching pictures (Dijkstra, Yaxley, Madden & Zwaan, 2004; Hirschfeld & Zwitserlood, 2010; Holt & Beilock, 2006; Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007; Madden & Dijkstra, 2010; Madden & Zwaan, 2006; Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). A symbolic theory of language comprehension, in which the eagle might be represented as a list of features or a node in a propositional network, cannot easily account for this mismatch effect. Instead, the mismatch effect suggests that meaning is instantiated by the partial reactivation and integration of previous perceptual experiences.

While the studies listed above typically involved undergraduates, or, in some cases, older adults, it is as of yet unclear whether children construct perceptual simulations during language comprehension. On the one hand, work such as Mandler's (1992; 2010) highlights the role of spatial representations in the development of conceptual knowledge. From there, it is a small step to proposing that perceptual simulations are central to young children's comprehension processes. On the other hand, the extent to which perceptual simulations are constructed as a function of language comprehension is constrained by domain expertise (e.g., Holt & Beilock, 2006) and processing capacity (e.g., Madden & Zwaan, 2006). These limitations may be especially restricting when children learn to read. This is because in the early stages of reading, processing resources are primarily allocated to breaking the orthographic code (Perfetti, 1985), rather than to mapping this code onto meaningful representations. We may therefore expect the use of perceptual simulations to follow different developmental trajectories for reading and listening. The present study explores this issue in 7 to 13-year-olds. This approach potentially informs theories of language development by showing how children represent meaning under different linguistic modalities, and at the same time feeds into theories of embodied cognition by showing to what extent developmental factors constrain the utility of perceptual simulations. The following sections will discuss in more detail the task of language comprehension and the role of domain expertise and processing efficiency.

Situation Models

It is a generally held view that when people comprehend language, they create a mental representation of the described state of affairs, rather than of the text itself (Johnson-Laird, 1983; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). This representation is referred to as a *situation model*. According to theories of embodied language comprehension, perceptual-motor simulations, not amodal propositions, are the building blocks of situation models. Readers and listeners construct these simulations by reactivating and integrating traces of previous experience distributed across multiple perceptual and motor modalities in the brain (Barsalou, 1999; Zwaan & Madden, 2005). The recruitment of perceptual and motor systems has been demonstrated in a host of behavioral tasks, showing that comprehenders do not only simulate the implied shape and orientation of objects, but many other perceptual features of a situation as well, such as an object's direction of motion (Glenberg & Kaschak, 2002; Kaschak et al., 2004; Zwaan, Madden, Yaxley, & Aveyard, 2004; Zwaan & Taylor, 2006), the rate and length of fictive motion (Matlock, 2004), the axis along which an action takes place (Richardson, Spivey, McRae, & Barsalou, 2003), the part of the visual field where a scene takes place (Bergen, Lindsay, Matlock, & Narayanan, 2007), the visibility of objects through an obscuring medium (Yaxley & Zwaan, 2007) and the sounds produced by entities (Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010). Efforts to build a situation model for larger stretches of discourse also engage visual processing mechanisms. For instance, memorizing a dot matrix for a later recognition test, which involves a visual-spatial load, has a stronger degrading effect on comprehension of short texts than remembering a letter string, which involves a verbal load (Fincher-Kiefer, 2001).

The mounting evidence for the automatic activation of experiential traces during language comprehension does not imply that these are the only means that individuals have at their disposal for representing meaning. One can achieve at least a rudimentary understanding of discourse by making use of the co-occurrences of linguistic forms. Latent Semantic Analysis (LSA; Landauer & Dumais, 1997) uses an algorithm that maps words into a high-dimensional semantic space. Linguistic forms (e.g., words, phrases or whole texts) are compared in this space, resulting in a cosine value representing their semantic relatedness. LSA does a remarkable job at simulating human performance, for instance by correctly answering multiple choice textbook questions after being trained on the content of the textbook (Landauer, Foltz, & Laham, 1998). Nevertheless, as long as linguistic forms, which are inherently abstract, merely refer to other linguistic forms, which are also inherently abstract, the semantic relatedness values are essentially meaningless. For a linguistic form to be meaningful, it must be grounded in experience outside the network (Glenberg & Robertson, 2000; Harnad, 1990; Searle, 1980). These concerns notwithstanding, linguistic representations may be functional to on-line language comprehension. This might work as follows: according to principles of content addressability and encoding specificity, the information in memory that is most similar to the cue becomes active most rapidly (e.g., Tulving & Thomson, 1973). When an incoming

word is recognized, activation spreads to associated linguistic representations, and subsequently to perceptual representations (see also Paivio, 1986). One possible function of the activation of linguistic representations is to anticipate upcoming information (e.g., Barsalou, Santos, Simmons, & Wilson, 2008; Zwaan, 2008). For example, upon hearing the word *bird*, the perceptual system is set up to activate a representation of not only *bird*, but also of *sky*, *fly* and *wings*, thereby facilitating activation and integration of these concepts in case any of these words will actually follow.

Whether linguistic or perceptual-motor representations dominate an individual's response on a given task depends on their domain-specific knowledge and processing capacity. Owing to the time-constrained nature of most comprehension tasks, it matters how quickly a perceptual representation is activated as a function of hearing or seeing a word and how quickly sentence context can be used to constrain the developing simulation (Madden & Zwaan, 2006). Moreover, although language affords the description and simulation of unfamiliar situations, possessing a rich network of experiential traces should facilitate the rapid activation of a trace that is appropriate in a given linguistic context. Importantly, both processing efficiency and domain-specific knowledge increase throughout childhood, but neither may be adequately developed in 7-year-olds to support the on-line construction of perceptual simulations. We turn to these issues below.

Domain Expertise

In many ways, development during childhood parallels that from novice to expert in specific domains. Holt and Beilock (2006) investigated whether domain expertise had a bearing on individuals' comprehension of descriptions of domain-specific situations. Novice and expert football and hockey players performed a sentence-picture verification task in the perceptual mismatch paradigm described above. When sentences dealt with everyday scenarios or actions anyone might perform (e.g., *The woman put the umbrella in the closet*), both groups performed accurately, making the correct decision on 96% of the trials. Also, both novice and expert athletes showed a mismatch effect (i.e., they responded more quickly to pictures that matched the previously presented sentence). However, when the sentences described sport-specific scenarios (e.g., *The coach saw the defenseman stop the kick*), both groups still performed accurately, but only expert athletes showed a mismatch effect. These findings suggest that possessing perceptual-motor representations depends on experience interacting with objects and performing the actions in question.

Analogously, adults and older children are more likely to possess rich networks of perceptual representations than younger children, even for everyday scenarios. This is reflected in the growth of the depth of vocabulary knowledge (e.g., Lahey, 1988; Ouellette, 2006). For example, a 13-year-old may have a multitude of perceptual traces associated with the word *pigeon*, having seen it in flight, walking, perched on a roof, having studied its colors in picture books, and having heard its cooing. This allows them

to readily use an appropriate trace for constructing a simulation of the sentence *Bob saw the pigeon in the sky*. Suppose a 7-year-old's only experience with a pigeon is seeing it perched, they may represent the flying pigeon by second-order grounding (Harnad, 1990), for instance by using their perceptual knowledge about other flying birds. In line with the findings by Holt and Beilock (2006), this way of grounding language in experience may not be adequately efficient to support time-constrained comprehension.

Processing Efficiency

Another crucial component of language comprehension is the ability to hold words and clauses in memory while processing other words and clauses until both can be integrated. This ability is measured by the reading span task (Daneman & Carpenter, 1980). Reading span is operationalized as the number of words that an individual can keep in memory while giving sensibility judgments about a set of unrelated sentences. As such, this measure directly taps into the efficiency of the comprehension process. Previous research has demonstrated that only high-span comprehenders immediately apply sentence-level context during comprehension (Madden & Zwaan, 2006; Van Petten, Weckerly, McIsaac, & Kutas, 1997). In one study, Madden and Zwaan (2006) compared high and low-span comprehenders on their performance on a sentence-picture verification task in the perceptual mismatch paradigm. The location was stated first, so that the target object was always the last word in the sentence (e.g., *In the pot there was spaghetti*). When the picture was presented 750 ms after the offset of the target word, both low and high-span comprehenders showed a mismatch effect. When the picture was presented immediately after the offset of the target word, only high-span comprehenders showed a mismatch effect. These findings suggest that high-span comprehenders were efficient at activating a contextually appropriate perceptual representation of the target word. Low-span comprehenders, on the other hand, were slower to construct a perceptual simulation and therefore relied on a linguistic representation.

Reading span increases during childhood (e.g., Case, Kurland, & Goldberg, 1981; Chiappe, Hasher, & Siegel, 2000), suggesting that older children use their available processing resources more efficiently than younger children. A possible explanation is that older children have developed stronger activation links between words and their associated perceptual representations. The reinforcement of such basic-level processes may improve the efficiency of the comprehension process as a whole (MacDonald & Christiansen, 2002). Accordingly, the ability to bring sentence context to bear on perceptual simulations may be limited at the age when children start learning to read, but considerably better developed after several years of education.

To summarize, relatively inefficient processing found in children and a lack of experience with the objects and actions being talked about may pose a threshold for constructing perceptual simulations during comprehension. In the present study, children from grades 2 to 6 (7 to 13-year olds) performed a picture verification task in the perceptual mismatch paradigm. The choice of these age groups is non-arbitrary with

regard to reading education. In grade 2, the focus shifts from decoding to comprehension. By the end of grade 6, formal reading instruction has typically finished, although development still continues after that. Hence, these age groups capture the stages during which the propensity to construct perceptual simulations during language comprehension is likely to change.

Experiment 1

Method

Participants. Children (N = 140, 62 boys) from grades 2 through 6 in an ethnically heterogeneous primary school in a large urban area in the Netherlands, participated in the study. Each grade contributed 28 children. Ages ranged from 7.5 years to 9.3 in grade 2 (mean 8.3), from 8.6 to 9.9 in grade 3 (mean 9.1), from 9.5 to 11.0 in grade 4 (mean 10.3), from 10.5 to 12.4 in grade 5 (mean 11.2), and from 11.7 to 13.3 in grade 6 (mean 12.3).

Participants were screened for abnormal comprehension by their teachers. This screening was complemented with standardized tests of comprehension, on which all participating children performed at age-appropriate levels. Caretakers were informed about the research and gave passive consent before the start of the experiment.

Materials. We constructed 42 experimental sentence pairs of the format '*Agent* saw the *object* in/on the *location*'. Each sentence implied a distinct shape (e.g., *Bob saw the pigeon in the nest* vs. *Bob saw the pigeon in the sky*) or orientation of the same object (e.g., *John saw the nail in the wall* vs. *John saw the nail in the floor*). The only difference within a given sentence pair was on the last noun and, in a few cases, the preposition. Some location nouns were used in two sentence pairs, but care was taken that subjects saw each location noun only once. All sentences were in Dutch. The sentences were pre-recorded by an adult male native speaker of Dutch and edited to terminate at the offset of the last word.

For each experimental sentence pair, we selected one picture of the described object. This picture matched the object's shape or orientation implied by the location in one of the sentences, but mismatched it in the other. The pictures were full-color photographs obtained from various web libraries and scaled to occupy an area of approximately 10 x 10 cm.

In addition to the experimental items, 76 filler items were constructed, 60 of which featured unrelated pictures and 16 of which featured a picture of the location. These served to balance the number of affirmative and negative responses and to prevent subjects from merely paying attention to the object noun. The total set of stimuli, intended to be used in both experiments reported here, consisted of 118 sentence-picture pairs, 58 requiring an affirmative response and 60 requiring a negative response.

To ensure that sentences were clearly audible and did not contain any unfamiliar words, we conducted a pilot study with students from grades 2, 3 and 6 (N = 12). No problems with sentences or individual words were reported. We also checked whether the

pictures were unambiguous and adequately fitted the described entities. Items that were responded to incorrectly by more than three participants were not included in the later experiment.

The picture verification and motor speed tasks were run using the E-Prime stimulus presentation software (Schneider, Eschman, & Zuccolotto, 2002). Responses were registered using a custom-made response box with two large buttons, 4 cm in diameter each, labeled *no* (left) and *yes* (right), approximately 20 cm apart from each other.

Procedure. The experiment took place in a quiet room within the school environment, where participants were seated in front of a computer screen. The experiment started with a simple button-press task, which used the same set-up as the picture verification task. This was because we expected substantial variance in response times due to differences in motor speed. To be able to correct for these differences, we measured the participants' reaction speed in the absence of higher-order cognitive processes. In a given trial, a cross appeared on the screen, either left- or right-aligned. Participants were asked to respond by pressing a button on the corresponding side of the response box. Participants rested their hands on the buttons and pressed as quickly as possible. Ten trials were presented in random order at an interval of 1000 ms. Only the five right-aligned trials were registered, since only the right hand would be used for giving correct responses in the picture verification task.

Subsequently, the experimenter explained participants they were about to listen to a set of sentences and that each time a picture would be shown afterwards. Participants were asked to rest their hands on the buttons labeled *yes* and *no* and to press as quickly as possible after they had determined if the depicted object had been mentioned in the sentence. They listened to the sentences and kept their eyes focused on a fixation point in the center of the screen. This point was replaced by the picture after 1000 ms following the offset of the sentence. Participants began with 10 practice trials, consisting of five related and five unrelated pictures. Next, they completed a sequence of 59 trials, including 21 experimental trials. Each participant performed 10 or 11 match trials and 10 or 11 mismatch trials, all requiring an affirmative response. In addition, there were eight fillers requiring an affirmative response, and 30 fillers requiring a negative response. All trials were presented in random order. The experiment took approximately 15 minutes to complete.

It is important to note that both experiments in the present study took place within a single session. No separate practice trials were offered in between. The order of the experiments was counterbalanced, so that half of the participants performed Experiment 2 before entering Experiment 1 and vice versa. Items were also counterbalanced across experiments, so that participants would not see the same picture in both experiments, and each picture occurred as often in Experiment 1 as it did in Experiment 2.

Results.

Preliminary analyses. Two experimental items were removed prior to the statistical analyses, owing to their large number of incorrect responses. Next, all trials with response latencies smaller than 300 ms or larger than 3000 ms were removed, which yielded an additional loss of less than 1% of the data. The average proportion of correct responses for all remaining trials, including fillers, was .96 (*SD* = .07). Filler items were not included in the reaction time analyses, but this check was performed to ensure that participants complied with the instructions and were not biased toward either affirmative or negative responses. The high percentage of correct responses indicates that participants adequately understood the procedure. The percentages of correct responses and average response times (collapsed over trials) for matching and mismatching experimental trials are given in Table 1. The percentages of correct responses were similar across conditions, indicating that mismatching trials were not more likely to elicit a negative response, warranting further comparison of response times.

Table 1
Accuracy and Response Times for Experimental Trials in Experiment 1

		Accuracy (%)		RT (ms)	
Grade	Condition	Mean	*SD*	Mean	*SD*
2	match	97.5	5.0	1265	518
	mismatch	95.0	9.1	1336	555
3	match	94.0	14.1	926	428
	mismatch	95.0	17.5	962	425
4	match	94.6	7.1	918	386
	mismatch	95.7	6.8	903	380
5	match	95.7	7.1	911	414
	mismatch	95.4	7.3	963	453
6	match	98.9	3.1	780	289
	mismatch	94.6	7.3	823	356

Analysis of response times. The common method of analysis used in response time research is to aggregate the observations over replications per participant. A disadvantage of this aggregation is that the sample size that is used is reduced to the number of participants in the sample, which reduces statistical power. As an alternative, a mixed model can be used in which random effects are included to model the dependencies between a participant's replications. We used multilevel analysis with items at the lowest and participants at the highest level. The dependent variable was response time and a random intercept was estimated to model the dependencies between trials within a participant.

In order to correct the effects of the variables of interest for other sources of variation in response time, we conducted a hierarchical regression analysis in which we first included the fixed- and random intercept and motor speed in the model. Table 2 shows the statistical tests for the mixed model effects. The random intercept was significant, indicating that after correcting for motor speed, the variance in response time between children was larger than zero. This justifies the inclusion of the random intercept in the model. Motor speed significantly predicted response times, $b = 1.43$, $SE = .12$, $r = .64$, $p < .001$.[1] Thus, the faster the performance in the motor speed task, the faster the performance in the picture verification task.

In the second step, the variables condition (match vs. mismatch), grade and Condition x Grade were added to the model. The main effect for condition was significant, $b = 81.50$, $SE = 30.81$, $r = .23$, $p = .008$. Corrected for grade and motor speed, responses to matching pictures were about 81 ms faster than responses to mismatching pictures. The main effect for grade was also significant, $Wald(4) = 61.86$, $p < .001$. Post-hoc comparisons with Bonferroni correction revealed a significant difference between grade 2 and 3 ($b = -247.58$, $SE = 48.04$, $r = .41$, $p < .001$). The negative coefficient indicates that children in grade 3 responded faster than children in grade 2, even after the correction for motor speed. Contrasts between other grades did not approach significance. Finally, the interaction effect for Condition x Grade did not approach statistical significance, $Wald(4) = 4.25$, $p = .37$. Thus, we found no evidence that the mismatch effect was influenced by grade.

Table 2
Hierarchical Regression Analysis of Response Times in Experiment 1

Block	Predictor	Wald *Z*	df	*p*
1	Intercept	1532.58	1	<.001
	Random Intercept	260.97	1	<.001
	Motor speed	139.19	1	<.001
2	Condition	7.00	1	.008
	Grade	61.86	4	<.001
	Condition x Grade	4.25	4	.37

Discussion

The main finding of Experiment 1 was that pictures were verified faster when they matched the preceding sentence than when they mismatched the preceding sentence.

[1] The effect size *r* was calculated as follows: $r = \sqrt{\frac{\left(\frac{b}{SE}\right)^2}{\left(\frac{b}{SE}\right)^2 + df}}$ (Rosnow & Rosenthal, 2005).

A straightforward interpretation of this finding is that participants had activated a mental representation of the target word that shared certain perceptual features with the picture probe. This primed the perceptual system with at least enough precision to speed up recognition of matching pictures relative to mismatching pictures (see also Hirschfeld & Zwitserlood, 2010). Crucially, the appropriate shape or orientation of the target word could only be derived by combining the object and the location to which the nouns in the sentence referred. For example, in the sentence *Martin saw the screw in the wall*, the horizontal orientation of the screw is not stated explicitly, but has to be inferred by meshing the affordances of a screw with those of a wall. Since neither a screw nor a wall does by itself enforce a horizontal representation, the mismatch effect must stem from perceptual simulation, and not from associative priming.

Grade had a substantial impact on response time, even when corrected for motor speed. Specifically, children in grade 3 responded faster than those in grade 2. Although we did not predict this increase in speed at this specific stage of development, it might be informative with respect to children's decision-making ability in language-based tasks. Alternatively, it is possible that the motor task did not capture all the variance related to motor speed, and that older children were faster simply due to more efficient response execution.

Importantly, although children in higher grades showed a nominally larger mismatch effect than children in lower grades, the interaction between condition and grade did not approach significance. So, whereas response latencies decreased with age, the additional time for verifying mismatching pictures remained constant. What explains this pattern? We speculate that the age-related variation in response time was distributed across recognition of the depicted object, accessing the name of that object, comparison of that name to the words in the sentence, and an affirmative or negative response based on that comparison. An alternative scenario, involving spillover processing of the sentence during the presentation of the picture probe for younger children, is less likely for two reasons. First, this would require the children to either process the sentence and the picture in parallel, or to suppress the picture until they had constructed a complete mental representation of the sentence. Given the saliency of the pictures, it is more likely that children immediately shifted their attention to the picture, performing the verification task using whatever representation of the sentence that was available to them. Second, if continuing processing of the sentence were the case, the children could have used the picture to aid their construction of a mental representation of the sentence. In that case, they would have exhibited less additional processing time for mismatching pictures.

Summarizing, Experiment 1 supports the notion that 7 to 13-year-olds simulate the implied shape and orientation of objects. The size of the mismatch effect did not increase as a function of grade, suggesting that even in grade 2, language about everyday situations is grounded in experience. With the present materials, we could not detect specifications beyond shape and orientation, so the possibility cannot be ruled out that older children formed richer situation models than younger children. Even so, this does

not undermine the notion that young children appear to activate and integrate perceptual memory traces while comprehending spoken language. Experiment 2 investigates whether the same holds for written language.

Experiment 2

Although there is evidence that a general cognitive mechanism underlies comprehension of spoken, written and even nonlinguistic information (e.g., Gernsbacher, 1985), we believe it is important to distinguish between written and spoken language, especially when reading is a newly acquired skill. Glenberg, Gutierrez, Levin, Japuntich and Kaschak (2004) point out that beginning readers often fail to map the words in written text to their referents, as opposed to words in speech. There are at least two reasons for this. First, spoken language is often used in highly determined contexts. When a child is first exposed to spoken language, there is a consistent, natural and repeated association between the words being uttered and the objects and events being referred to (Masur, 1997). For example, a caregiver may talk about 'the bottle' while holding a bottle, or say 'wave bye-bye' while actually waving (Glenberg et al., 2004). When a child learns to read, this association is broken. Written language often deals with objects and events outside of the reader's physical environment, so their referents need to be retrieved from memory. Second, when children have to read text themselves, their attention may be chiefly directed towards getting the spelling-to-sound conversions right, instead of retrieving and integrating the appropriate meaning representations (Perfetti, 1985). This allocation of resources is formalized in the Reader model (Just & Carpenter, 1992), according to which readers prioritize basic processes at the cost of higher-order comprehension processes. In line with this, there are numerous studies showing that less-skilled word reading is associated with poor comprehension (e.g., Muter, Hulme, Snowling, & Stevenson, 2004; Perfetti & Hart, 2001; Shankweiler, 1989).

From these considerations, it follows that even if children are proficient at constructing perceptual simulations in the domain of oral language, this is not necessarily true for the domain of written language. If less-skilled word readers do not reliably activate a perceptual representation as a function of seeing a word, or are compromised on the processing resources needed for using the linguistic context, we expect them not to show a mismatch effect for sentences they have to read themselves. This was investigated in Experiment 2, in which we compared participants of high and low word reading ability on their performance on a sentence-picture verification task.

Method

Participants. The participants in this experiment were children from grade 2 through 4 (N = 78, 38 boys) who also participated in Experiment 1. There were 27 children from grade 2 (mean age 8.2, ranging 7.5 to 9.3), 28 children from grade 3 (mean age 9.1, ranging 8.6 to 9.9), and 23 children from grade 4 (mean age 10.3, ranging 9.5 to 11.0).

Participants were assigned to groups according to their word reading ability scores. Word reading ability was measured with the *Three-Minute-Test* (TMT; Verhoeven, 1995), a standardized test which takes into account both speed and accuracy. Children read words from three lists of increasing complexity (monosyllabic words with single consonants, monosyllabic words with consonant clusters, polysyllabic words). Each list was shown for one minute. The score was calculated by subtracting the number of incorrectly pronounced words from the number of correctly pronounced words. Less-skilled readers (n = 20, mean age 8.4, ranging 7.6 to 9.4) had TMT scores of 55 or lower (M = 44, SD = 7). Skilled readers (n = 58, mean age 9.4, ranging 7.5 to 11.0) had TMT scores of 56 or higher (M = 76, SD = 12). This split corresponds to the norm of minimally acceptable word reading ability in grade 2. As such, less-skilled word readers in our sample can be assumed to be representative of less-skilled word readers in the population in general.

Materials. The sentence-picture pairs were identical to Experiment 1, except that sentences were presented as text on screen. The sentences were centered and displayed in a black 18-point Courier New font against a white background. To ensure that the sentences did not surpass the children's word reading ability, we conducted a pilot study with students from grades 2, 3 and 6 (N = 12). There were no problems with the object and location nouns in the experimental sentences. However, some names were pronounced incorrectly by one or more children. These names were replaced, along with the incorrectly pronounced object nouns in filler sentences.

Procedure. Participants sat in front of a computer screen and were instructed to read the sentences aloud. Incorrectly pronounced or skipped words were recorded by the experimenter. At the offset of the last word, the experimenter immediately pressed a button to replace the sentence by a fixation point. After 1000 ms, this point was replaced by the picture probe.

Half of the participants had not completed Experiment 1 before and started with 10 training trials. All participants then performed a sequence of 59 trials, including 21 experimental trials. The experimental trials consisted of 10 or 11 match trials and 10 or 11 mismatch trials, all requiring an affirmative response. In addition, there were eight fillers requiring an affirmative response, and 30 fillers requiring a negative response. The trials were presented in random order. The experiment took approximately 15 minutes to complete.

Results

Preliminary analyses. Trials that contained incorrectly pronounced or skipped words (1.4% of all trials) were removed from the data set. The two low-accuracy pictures from Experiment 1 also figured in Experiment 2 and were also removed. Next, all trials with response times greater than 3000 ms were eliminated, yielding an additional loss of less than 1% of data. The average proportion of correct responses for all remaining trials, including fillers, was .95 (SD = .08). Filler items were not included in the reaction time

analyses, but this check was performed to ensure that participants were not biased toward either affirmative or negative responses. The high percentage of correct responses indicates that participants adequately understood the procedure. The percentages of correct responses and average response times (collapsed over trials) for matching and mismatching experimental trials are given in Table 3. The percentages of correct responses were similar across conditions, indicating that mismatching trials were not more likely to elicit a negative response, warranting further comparison of response times.

Word reading and grade were significantly correlated (r = .70, p < .001). To rule out potential multicollinearity issues, we checked the unique contributions of both predictors by computing their partial correlations. Controlling for word reading, grade and response time were significantly correlated (r = -.19, p < .001). Controlling for grade, word reading and response time were significantly correlated (r = -.13, p < .001). We concluded that both variables accounted for unique portions of variance in response times. Hence, both variables were included in subsequent analyses.

Table 3
Accuracy and Response Times in Experiment 2

		Accuracy (%)		RT (ms)	
Word Reading	Condition	Mean	*SD*	Mean	*SD*
Low (n = 20)	match	94.7	10.9	1286	513
	mismatch	93.5	15.4	1375	515
High (n = 58)	match	96.9	5.7	979	387
	mismatch	95.4	6.4	1037	430

Analyses of response times. We used multilevel analysis with items at the lowest and participants at the highest level. As in Experiment 1, we conducted a hierarchical regression analysis in which we included the fixed- and random intercept and motor speed first. Table 4 shows the statistical tests for the mixed model effects. The random intercept was significant, which justifies the inclusion of a random intercept in the model. Motor speed significantly predicted response times (b = 1.95, SE = .19, r = .71, p < .001). Thus, the faster the response on the motor speed task, the faster the response on the picture verification task.

In a second step, the following block of predictors was entered: condition (match vs. mismatch), word reading (high vs. low), grade, Condition x Word Reading and Condition x Grade. The main effect for condition was significant, b = 79.98, SE = 36.19, r = .26, p = .027. Correcting for grade, word reading and motor speed, responses to matching pictures were about 80 ms faster than responses to mismatching pictures. The main effect for word reading did not approach significance, b = -53.31, SE = 80.14, r = .08, p = .51. This means that we found no difference between skilled and less-skilled word

readers in response time when controlling for grade and motor speed.[2] The main effect for grade was significant, $Wald(2) = 27.11$ $p < .001$. Post-hoc comparisons with Bonferroni correction applied revealed a significant difference between grade 2 and 3, $b = -368.16$, $SE = 100.20$, $r = .40$, $p < .001$. The negative coefficient indicates that children in higher grades responded faster than children in lower grades, even after correcting for motor speed and word reading skill. The difference between grade 3 and 4 did not approach statistical significance.

The interaction effect for Condition x Word Reading was not significant, $b = 28.82$, $SE = 71.10$, $r = .05$, $p = .69$. This means that word reading ability had no bearing on the size of the mismatch effect. Finally, the interaction effect for Condition x Grade did not approach significance, $Wald(2) = 1.22$, $p = .54$. Thus, there is no indication that the mismatch effect changed as a function of grade.

Table 4
Hierarchical Regression Analysis of Response Times in Experiment 2

Block	Predictor	Wald *Z*	df	*p*
1	Intercept	1648.35	1	<.001
	Random Intercept	100.85	1	<.001
	Motor Speed	201.99	1	<.001
2	Condition	4.88	1	.027
	Word Reading	.44	1	.510
	Grade	27.11	2	<.001
	Condition x Word Reading	.16	1	.690
	Condition x Grade	1.22	2	.540

Discussion

In Experiment 2, a sentence-picture mismatch effect was observed for written language. The interaction between condition and word reading skill did not approach statistical significance, indicating that skilled and less-skilled word readers showed a similar advantage for matching pictures. Although we anticipated that non-fluent word reading would interfere with the process of retrieving and integrating the appropriate memory traces, it did not preclude a perceptual simulation of the described situation.

It should be noted that children did not read the sentences silently, but aloud. It has been found that the performance aspect of reading aloud in the presence of an audience, for instance an experimenter, hampers comprehension, relative to reading aloud to oneself or reading silently (Holmes, 1985). If anything, this should decrease the

[2] Alternatively, word reading scores could be treated as a continuous variable. Because using raw word reading scores did not change the pattern of results, we use two discrete groups for ease of interpretation.

likelihood of perceptual simulation taking place, making the observed mismatch effect even more surprising.

One possible explanation for not finding larger effects related to word reading is that the words and syntactic constructions used in the sentences were too easy for individual differences in word reading ability to emerge. However, the reading times per sentence were longer for the less-skilled readers (mean 5.0 versus 3.6 sec), suggesting that their performance was well below ceiling.

Parallel to Experiment 1, a main effect was observed for grade. Older children responded faster than younger children, even when motor speed was partialed out. The decrease in response times was most notable between grade 2 and grade 3. Again, it is not likely that this difference is explained by additional time needed to process the sentence alone. Rather, it is the sum of the time needed to recognize the picture, access the name of the picture, compare the name to the words in the sentence, and the actual response based on that comparison. This would be consistent with the previous finding that less-skilled readers show impaired performance relative to skilled readers on tasks that require explicit comparison between a test probe and the preceding context (Long, Seely, & Oppy, 1999).

General Discussion

Two experiments addressed the question of whether 7 to 13-year-olds construct perceptual simulations during language comprehension. The results suggest that they do, while listening to sentences in Experiment 1, and while reading sentences aloud in Experiment 2. The children's responses in a sentence-picture verification task were consistent with the hypothesis that they had formed a perceptual simulation of the described situation, which they had constructed within 1000 ms after the offset of the critical location noun. Although response times were consistently longer after written sentences than after spoken sentences, the mismatch effect was comparable in magnitude across experiments, and also to that obtained with adults with similar procedures (e.g., Stanfield & Zwaan, 2001; Zwaan, Stanfield & Yaxley, 2002). Moreover, the mismatch effect emerged as a robust phenomenon, as it did not increase as a function of grade or word reading skill. This is surprising, given earlier demonstrations of the constraining role of expertise (Holt & Beilock, 2006) and processing capacity (Madden & Zwaan, 2006). Before discussing the implications of these findings for the development of language comprehension, it is important to rule out the possibility that children constructed perceptual simulations solely as a function of the given task. There are at least three ways in which this might be possible, but none of these possibilities appears to hold.

First, the experiments might have provided the opportunity for drawing backward inferences. That is, the shape of the described object might have been inferred only after viewing the picture probe. This interpretation of the data, however, runs counter to what the mismatch appears to be, namely an effect on picture recognition (see also Hirschfeld & Zwitserlood, 2010), not on the comparison between the name of the picture and the representation of the sentence. In fact, it is difficult to conceive of a locus of the mismatch

effect beyond lexical access to the name of the picture. After all, the name for the picture, on which the comparison is based, was the same in both conditions.

Second, the effects might be attributed to participants purposefully constructing mental images. Evidence from two previous studies speaks against this possibility. Pecher, van Dantzig, Zwaan and Zeelenberg (2009) showed that matching sentences facilitated picture recognition more than mismatching sentences when the recognition task was administered 45 minutes later, unexpectedly, following an unrelated filler task. In a similar vein, Wassenburg and Zwaan (2010) found longer reading times for mismatching sentences, 20 minutes after participants had viewed a set of pictures, being fully unaware of their relevance to a later reading task. Although the findings from these studies cannot be held conclusive for the child population in our sample, they provide compelling evidence that language comprehenders retain the shape and orientation of objects.

Third, it might be the case that the involvement of pictures modified the processes operating during or after the sentence presentation. Louwerse and Jeuniaux (2010) showed that the extent to which perceptual representations govern an individual's response on a given task, depends on both the instructions and the stimuli that are used. Participants in their study saw pairs of words or pictures that were presented in an iconic (e.g., *attic* above *basement*) or reverse-iconic vertical configuration (e.g., *basement* above *attic*), and were asked to judge their semantic relatedness or their iconicity. When making semantic relatedness judgments and seeing words as stimuli, response times and error rates were explained best by the order in which the words most frequently occur in language use. When making iconicity judgments or seeing pictures as stimuli, response times and error rates were explained best by the iconicity of the word pair's configuration. One might argue that the use of pictures in the present experiments favored perceptual representations over linguistic representations. However, participants were not asked whether the depicted object provided a good fit to the sentence they read before, but simply whether this object had been mentioned in the sentence. Thus, the use of perceptual representations does not seem to be strongly encouraged in the present experiments. To put our conclusions on more solid footing, future research could investigate whether similar effects can be obtained with verbal-only materials that, in addition, do not involve a judgment task. Nonetheless, there is good reason to believe that the construction of perceptual simulations was spontaneous, and did not reflect task-specific processing strategies.

We are now in a position to further discuss the relevance of our findings for developmental theory. An important finding was the absence of an interaction between condition and grade. Even the youngest children showed a mismatch effect, and its size did not increase for children in higher grades. This suggests that the tendency to form perceptual simulations for comprehending sentences such as *Bob saw the pigeon in the nest* and *Bob saw the pigeon in the sky* in perceptual experience is present by the time children enter second grade. At the same time, response latencies sharply decreased after grade 2, indicating that older children are more efficient at performing the task. The lack

of an interaction between condition and grade in conjunction with a main effect of condition may be difficult to reconcile with the view that meaning is represented by abstract symbols that are enriched by embodied representations when these become more easily available through experience. Rather, children construct simulations of events even if the knowledge of the objects involved is presumably limited.

Another crucial factor constraining the use of perceptual simulations, according to our discussion of the literature, was hypothesized to be processing efficiency (cf. Madden & Zwaan, 2006). This would be reflected in different developmental trajectories for listening and reading, given that the children's processing resources would be compromised by the reading task. Only with more fluent word reading should the effects in the reading experiment align with those in the listening experiment. An implication of this would be that for children of the same age, listening leads to more perceptual-like situation models than reading. However, the data clearly suggest a different state of affairs. It appears that perceptual simulations are used even when the efficiency of the linguistic processes giving rise to them is still developing. Consistent with this, children as young as 4 years have been found to mentally represent the spatial perspective of characters in narratives (Rall & Harris, 2000; Ziegler, Mitchell, & Currie, 2005), as well as their movement (Fecica & O'Neill, 2010).

Importantly, the present results should not be taken as evidence that the construction of perceptual simulations is equally efficient across all grades studied. In both experiments, participants had 1000 ms to construct a perceptual simulation on the basis of the linguistic input before viewing the picture. Future research could investigate whether shorter intervals are informative as to the time course of the activation and integration of perceptual representations as a function of age.

Overall, the results suggest that perceptual simulations are constructed even when expertise and processing capacity are relatively limited. Perceptual simulation may play a more important role in developing language comprehension than previously thought. This is the first study to directly address this issue (but see Glenberg et al., 2004, for a comparable discussion). Although no detailed framework as of yet exists in the literature, certain accounts may accommodate these findings. For instance, our interpretation of the results is broadly consistent with theories that place the manipulation of spatial representations at the core of developing cognition (e.g., Mandler, 2010). Under such a view, it is plausible that thought involves the manipulation of perceptual rather than abstract symbols, and that language is one of the mechanisms that drive this manipulation. Learning to comprehend language, then, means learning how to use language as a tool for evoking appropriate simulations.

Finally, while our results are supportive of theories of embodied language comprehension, several recent papers have outlined fundamental challenges for such theories (Mahon & Caramazza, 2008; Zwaan, 2009). In particular, the field is in need of research that shows whether embodied representations are essential to comprehension, or are epiphenomenal to other processes. While underscoring the importance of addressing

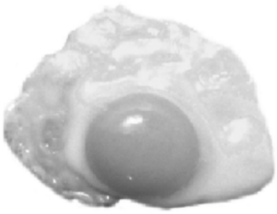
these challenges, we believe that the present work holds value in that it extends the descriptive and explanatory power of the simulation view of language comprehension to development during childhood, for both spoken and written materials, and for skilled and less-skilled word readers alike.

Appendix A

Samples of Experimental Sentence-Picture Pairs

Martin saw the screw in the wall/ceiling

Steve saw the toothbrush in the cup/sink



Luke saw the egg in the skillet/box

Bob saw the pigeon in the sky/nest

Appendix B

List of Experimental Sentences (in Dutch) and their English Translations

Jeroen zag de banaan op het toetje	Jerome saw the banana on the dessert
Jeroen zag de banaan aan de tros	Jerome saw the banana on the bunch
Laura zag de ananas op de taart	Laura saw the pineapple on the pie
Laura zag de ananas in de tas	Laura saw the pineapple in the bag
Jan zag de appel in de fruitsla	John saw the apple in the fruit salad
Jan zag de appel in de rugzak	John saw the apple in the backpack
Tom zag de appeltaart op zijn bordje	Tom saw the apple pie on his plate
Tom zag de appeltaart in de oven	Tom saw the apple pie in the oven
Moeder zag de kip in de oven	Mother saw the chicken in the oven
Moeder zag de kip in het hok	Mother saw the chicken in the coop
Jens zag de hond in de mand	James saw the dog in the basket
Jens zag de hond op het grasveld	James saw the dog on the lawn
Vera zag de kat op het hek	Vera saw the cat on the fence
Vera zag de kat in de mand	Vera saw the cat in the basket
Kevin zag het brood in de zak	Kevin saw the bread in the bag
Kevin zag het brood op zijn bord	Kevin saw the bread on his plate
Simon zag het vlees op zijn bord	Simon saw the meat on his plate
Simon zag het vlees in de koelkast	Simon saw the meat in the fridge
Rob zag de duif in het nest	Bob saw the pigeon in the nest
Rob zag de duif in de lucht	Bob saw the pigeon in the sky
Fleur zag de duiker in het water	Flora saw the diver in the water
Fleur zag de duiker op de kant	Flora saw the diver on the beach
Karel zag de eend in de lucht	Carl saw the duck in the sky
Karel zag de eend in de vijver	Carl saw the duck in the pond
Luuk zag het ei in de doos	Luke saw the egg in the box
Luuk zag het ei in de pan	Luke saw the egg in the skillet
Bert zag de spaghetti in de pan	Bert saw the spaghetti in the pot
Bert zag de spaghetti in de verpakking	Bert saw the spaghetti in the wrapping
Ruud zag de danseres in de kleedkamer	Ray saw the dancer in the dressing room
Ruud zag de danseres op het podium	Ray saw the dancer on the stage
Kim zag de handdoek op de plank	Kim saw the towel on the shelf
Kim zag de handdoek aan het haakje	Kim saw the towel on the hook
Liza zag de kaas in de muizenval	Lisa saw the cheese in the mouse trap
Liza zag de kaas op het broodje	Lisa saw the cheese on the bun
Alex zag de ballon in het zakje	Alex saw the balloon in the wrapper
Alex zag de ballon aan het touwtje	Alex saw the balloon on the string
Martijn zag de bloes op de hanger	Martin saw the shirt on the hanger
Martijn zag de bloes op de stapel	Martin saw the shirt on the pile
Kees zag de meloen in de tuin	Keith saw the melon in the garden
Kees zag de meloen in de kom	Keith saw the melon in the bowl
Dirk zag de parasol in de kelder	Dick saw the parasol in the cellar
Dirk zag de parasol op het terras	Dick saw the parasol on the terrace
Vader zag de sigaret in het pakje	Father saw the cigarette in the pack
Vader zag de sigaret in de asbak	Father saw the cigarette in the ashtray

Bas zag de citroen in het drankje	Barry saw the lemon in the drink
Bas zag de citroen op de fruitschaal	Barry saw the lemon in the drink
Noa zag de tennisster op de stoel	Noah saw the tennis player on the chair
Noa zag de tennisster op de baan	Noah saw the tennis player on the court
Eva zag de tomaat aan de tros	Eve saw the tomato on the bunch
Eva zag de tomaat op de pizza	Eve saw the tomato on the pizza
Lieke zag de ui in de tas	Lea saw the onion in the bag
Lieke zag de ui op de hamburger	Lea saw the onion on the hamburger
Fred zag de auto op de racebaan	Fred saw the car on the race track
Fred zag de auto in de garage	Fred saw the car in the garage
Iris zag de voetballer op het bankje	Iris saw the football player on the bench
Iris zag de voetballer op het veld	Iris saw the football player on the pitch
Stef zag de tandenborstel in de beker	Steve saw the toothbrush in the cup
Stef zag de tandenborstel op de wastafel	Steve saw the toothbrush on the sink
Ruben zag de tak op de stoep	Ruben saw the branch on the curb
Ruben zag de tak in de grond	Ruben saw the branch in the ground
Maarten zag de schroef in het plafond	Martin saw the screw in the ceiling
Maarten zag de schroef in de muur	Martin saw the screw in the wall
Wouter zag de lamp aan de muur	Walt saw the lamp on the wall
Wouter zag de lamp aan het plafond	Walt saw the lamp on the ceiling
Bart zag de spijker in de vloer	Bart saw the nail in the floor
Bart zag de spijker in het prikbord	Bart saw the nail in the bulletin board
Leo zag de dartpijl in de roos	Leo saw the dart in the bullseye
Leo zag de dartpijl in de vloer	Leo saw the dart in the floor
Moeder zag de vork in de la	Mother saw the fork in the drawer
Moeder zag de vork in de biefstuk	Mother saw the fork in the steak
Roos zag de lepel in de soepkom	Rose saw the spoon in the soup bowl
Roos zag de lepel op het tafelkleed	Rose saw the spoon on the tablecloth
Hugo zag de veer in de inktpot	Hugo saw the quill in the ink jar
Hugo zag de veer in het kippenhok	Hugo saw the quill in the chicken pen
Sara zag de sleutel in het slot	Sara saw the key in the lock
Sara zag de sleutel aan het haakje	Sara saw the key on the hook
Daan zag het boek op de leestafel	Danny saw the book on the reading table
Daan zag het boek in de kast	Danny saw the book in the cabinet
Emma zag de rits aan de tas	Emma saw the zipper on the bag
Emma zag de rits aan het vest	Emma saw the zipper on the waistcoat
Vader zag de pet aan de kapstok	Father saw the cap on the peg
Vader zag de pet op het hoofd	Father saw the cap on the head
Gijs zag de boom op de heuvel	Gus saw the tree on the hill
Gijs zag de boom op de vrachtwagen	Gus saw the tree on the truck

Chapter 3

Eye Movements Reveal Differences in Children's Referential Processing During Narrative Comprehension*

* This chapter has been published as Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2014). Eye movements reveal individual differences in children's referential processing during narrative comprehension. *Journal of Experimental Child Psychology*, *118*, 57-77.

Abstract

Children differ in their ability to build referentially coherent discourse representations. Using a visual world paradigm, we investigated how these differences might emerge during the online processing of spoken discourse. We recorded eye movements of 69 children (ages 6-11) as they listened to a 7-minute story and concurrently viewed a display containing line drawings of the protagonists. Throughout the story, the protagonists were referenced either by a name (e.g., "rabbit") or an anaphoric pronoun (e.g., "he"). Results showed that the probability of on-target fixations increased after children heard a proper name, but not after anaphoric pronouns. However, differences in the probability of on-target fixation at word onset indicate that the referents of anaphoric pronouns were anticipated by good comprehenders, but less so by poor comprehenders. These findings suggest that comprehension outcomes are related to the online processing of discourse-level cues that regulate the accessibility of entities.

When comprehending a story, it is important to keep track of who does what to whom. Many narrative texts, however, are not entirely explicit about this. Take the sequence *John decided to call Bill. He still owed him a favor.* Interpretations of John owing Bill and Bill owing John are both plausible. Because for sequences like this there is a statistical tendency for the grammatical subject of the second clause to refer back to the first-mentioned entity of the first clause (Gernsbacher & Hargreaves, 1988), most listeners will go with the former interpretation. This discourse-level cue to reference requires extensive exposure to be learned. Children at the age of 5 use the gender of pronouns (Arnold, Brown-Schmidt, & Trueswell, 2007) and the semantics of verbs (Pyykkönen, Matthews, & Järvikivi, 2010) to guide referential choice, but they do not yet show a first-mention bias, neither in offline nor in online tasks (Arnold et al., 2007). Given this protracted developmental trajectory, there may be much interindividual variation in the ease with which children process anaphoric pronouns (i.e., pronouns that refer to an entity introduced earlier in the discourse, called an antecedent) in the next years of life. Indeed, Yuill and Oakhill (1988) cite two studies reporting 77% correct resolution in normal 9-year-old readers (Bormuth, Manning, Carr, & Pearson, 1970), and 57% in poor readers after training in anaphor resolution (Dommes, Gersten, & Carnine, 1984), while their own data showed performance of 91% for 8-year-old good readers and 66% for poor readers, even when the pronoun and the antecedent were close together in the text.

In this paper we explore this state of affairs as a source of differential comprehension outcomes in school-aged children. To what extent does understanding a story depend on accurate online referential processing? This question fits in with the increasing tendency to define skilled and struggling developing comprehenders not only by their offline performance on a particular task, but also by the online processes that lead up to it (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007). While various frameworks of text comprehension underline the importance of referential coherence (e.g., Gernsbacher, 1990; Myers, O'Brien, Albrecht, & Mason, 1994; Sanford & Garrod, 1981; van den Broek, Young, Tzeng, & Linderholm, 1999; Zwaan & Radvansky, 1998), online referential processing has not been specifically related to understanding of extensive discourse.

In the remainder of the introduction, we will discuss what it means to successfully comprehend discourse and how readers construct referential coherence. We pay specific attention to the notion that different types of referring expressions, while they may be referentially equivalent, have different functions in organizing the flow of information within the discourse. We then place these aspects in a developmental context and propose a novel application of the visual world paradigm to investigate them in a way that is temporally sensitive and unobtrusive to the comprehension process.

Referential Coherence and Discourse Comprehension

Narrative comprehension is widely assumed to entail the construction of a situation model, an integrated representation of the events described by the text (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Narratives typically revolve around a limited number of characters, whose goals and plans are the driving force behind the events that take place. For that reason, protagonists have been called the 'meat' of situation models (Zwaan & Radvansky, 1998). When a protagonist is first introduced in a story, usually by a name, comprehenders set up a mental representation. As subsequent events unfold, a major task for the listener or reader is to identify whom a given sentence is about, and to connect the incoming information with what he or she already knows about this character (Morrow, 1985). To determine whether incoming information coheres with previously comprehended information, readers and listeners use various linguistic cues (Givón, 1992). According to Gernsbacher (1997), these cues lie on a continuum from explicit to implicit. Whereas a repeated name or a noun with a definite article (e.g., *John* or *the old man*) can be mapped onto the existing representation of a discourse entity in a relatively straightforward manner, a pronoun (e.g., *he*) has no meaning outside the scope of its immediate context and requires knowledge-driven inferential processing. There is evidence that resolving anaphoric pronouns that require an inference is problematic for children as old as 5 years (e.g., Wykes, 1981, 1983). In an acting-out task, children made more mistakes in reproducing the second sentence of a pair when it contained pronouns (e.g., *Jane needed Susan's pencil. She gave it to her*) than when it contained nouns and proper names (e.g., *Susan gave the pencil to Jane*). In this case surface-level cues, such as grammatical subjecthood and order-of-mention, are not diagnostic and applying knowledge about how events in the world are related is the only way to arrive at the correct interpretation.

It is important to note, however, that in most spoken and written discourse, names and pronouns are not used as interchangeably as the example above might suggest. An anaphoric pronoun is mostly used to signal referential continuity, while a name or a noun phrase indicates a shift from one entity to another. Accordingly, the mental operation cued by a pronoun is to maintain the currently activated entity in working memory, or to maintain one of a larger set of currently activated entities while suppressing the others. By contrast, a name is often used to move the attentional focus from one entity to another and will most likely trigger a search for the appropriate entity in episodic memory (Chafe, 1994; Givón, 1992). The expectation on the part of readers and listeners for a name to refer to a less accessible entity is reflected in the 'repeated-name penalty': when a sentence contains a name that refers back to a highly accessible entity, reading times for that sentence are longer than when it contains a pronoun (Gordon, Grosz, & Gilliom, 1993).

The choice for a more or less specified referring expression is not only determined by what has been accessible before ('looking backward'), but also by what is likely to continue to be important to the speaker ('looking forward'; Arnold et al., 2007; Grosz, Joshi, & Weinstein, 1995). Applied to the example above, this means that comprehenders would come to a third sentence with different expectations and accessibility levels regarding the characters after *Susan gave the pencil to Jane* than after *Susan gave it to her*. In sum, names and anaphoric pronouns do not only implicate a different load on inferential processing, they also occur in different contexts and have different functions in directing the reader's or listener's attention. Therefore, to adequately process a referring expression does not only mean to tie it to the mental representation of the appropriate entity, but also to use these properties in managing the flow of linguistic information.

Development of Referential Processing

How do school-aged children fare at processing anaphora for constructing referential coherence? Similar to the results reported by Wykes (1981), Oakhill and Yuill (1986) found that resolving anaphora that required an inference was problematic for a subgroup of 8-year-olds, especially when the inference was complex. For instance, providing the appropriate pronoun in the sentence *Steven gave his umbrella to Penny in the park because … wanted to keep dry*, which involves the chain of inferences that a person who wants to keep dry is likely to want an umbrella, and is therefore also its likely recipient, yielded an error rate of 17% in good comprehenders, and 33% in poor comprehenders. An analogous pattern of errors was found when children had to indicate whether a statement following the sentence was true or false. Although the link with anaphor resolution has not been directly tested, research suggests that difficulties in inference generation cannot be solely attributed to weak retention of literal information (e.g., Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 1999; Omanson, Warren, & Trabasso, 1978).

Another source of difficulty is the distance between an anaphoric device and its antecedent. Moberly (1978) had 9-year olds read brief stories, which were replaced by a new copy in which the anaphors were underlined. Children were asked to write down the words that each of these anaphors pointed back to. The number of errors increased as the textual distance between the two elements increased. Also, poor comprehenders' performance was affected more by textual distance than good comprehenders'. Yuill and Oakhill (1988) used an oral version of the task with a single story, where the experimenter stopped to ask the child what each anaphor referred to, and, in case of an incorrect answer, posed an additional textual question that relied on correct resolution (e.g., *Who carried his rod to the bus stop?*). They obtained similar results with respect to the difference between good and poor comprehenders. Also, poor comprehenders gave more wrong answers

(rather than no answers) to the textual questions, suggesting that their incorrect resolution of anaphora led to misinterpretations of the text.

A parallel may be drawn between the construction of referential coherence and the development of other components of comprehension skill. If resolving anaphora is contingent upon making inferences, then the difficulties young children experience in other inferential tasks should abide by the same constraints. Indeed, research in other paradigms shows that 6-year-olds can infer causal connections between physical events that are adjacent in a text (Casteel, 1993), but often fail to do so for events that are further apart (Ackerman & McGraw, 1991) or involve abstract, non-physical relations, such as a character's goals or emotions (van den Broek, Lorch, & Thurlow, 1996). By the age of 11, these causal and goal-related inferencing abilities are usually well developed, and similar to those of adults (see van den Broek, 1997, for a review), although substantial individual differences continue to exist (e.g., Cain & Oakhill, 1999; Rapp et al., 2007). A crucial difference, however, is that the construction of referential coherence is facilitated by additional syntactic cues, such as the gender and number of a pronoun, and distributional cues, such as order-of-mention, which may lessen the need for effortful inferential processing. Proficiency in using these types of cues may develop independently from inference-making.

Before outlining this development, we should point out that the studies discussed above employed fairly explicit measures of pronoun understanding. As a result, the data may not reflect routine comprehension processes, in that children do not necessarily pay the same attention to resolving anaphors during normal reading. Also, the tasks rely to some extent on metalinguistic awareness, which may constitute a bottleneck for children with low verbal skills. Moreover, the methods provide little in the way of temporal resolution. Knowing the time course of comprehenders' interpretation of anaphora can provide valuable additional information about the underlying processes.

Referential Processing in the Visual World Paradigm

There are several studies that employed the visual world paradigm to obtain an online measure of the resolution of anaphora, such as pronouns (e.g., *he/she*; Arnold et al., 2007; Song & Fisher, 2005), demonstratives, (e.g., *this/that*; Kaiser & Trueswell, 2008) and reflexives (e.g., *himself/herself*; Clackson, Felser, & Clahsen, 2011; Kaiser, Runner, Sussman, & Tanenhaus, 2009), which all derive their meaning from the local discourse context. The picture that emerges from this work is that across development, context has a differential effect on how individuals resolve pronouns. For instance, the first-mention heuristic is an effective cue for adults to determine the referent of a pronoun. With no other cues present, adults make a saccadic eye movement toward the first-mentioned entity approximately 200 ms after the onset of the pronoun (e.g., a picture of Donald Duck while hearing *Donald is bringing some mail to Mickey, while a violent storm is beginning.*

He's carrying an umbrella, and it looks like they're both going to need it; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000). This is not the case for 4 and 5-year-old children, who are equally likely to look at the first- and the second-mentioned entity (Arnold et al., 2007). However, when the accessibility of the first-mentioned entity is enhanced by repeating it in two consecutive sentences prior to the target sentence, children do show a preference for that entity, but only between 1 and 2 sec after the onset of the pronoun (Song & Fisher, 2005). Thus, both sensitivity to discourse constraints and the ability to rapidly use these constraints develops after early childhood.

To summarize, findings from a variety of experimental methods suggest that the interpretation of discourse anaphora undergoes development well into childhood, and that there are considerable individual differences between children of the same age. Yet, little is known about the trajectory of this development during the school years, other than from the offline tasks discussed above. These indicate that inferencing ability and text characteristics likely play a role, but it is not clear how both affect children's real-time consideration of discourse referents. Furthermore, while a number of visual world studies have shed light on the time course of referential solution in different stages of cognitive development, their generalizability may be hampered by two factors. First, the target sentence is often the last or second last in a sequence, with the visual display remaining visible for several seconds afterwards. Such an interval in which the auditory stimulus can be processed without competition from incoming information is not representative of most spoken discourse, where utterances usually follow each other in quick succession. Consequently, it might be that children's ability to interpret pronouns, if not overestimated, is viewed under conditions that do not accurately reflect the demands of normal discourse comprehension. Second, new protagonists and scenes are usually introduced for each measurement. While this is clearly justified by the need for a controlled experimental design, it may induce a bias for characters that are still relatively new and unfamiliar, leaving open the question of how comprehenders process referring expressions in a more extensive discourse context.

To overcome these limitations , we aimed to more closely match the task demands of discourse comprehension by implementing a longer story into a visual world setting. Why has such an approach not been taken before? This is perhaps not surprising. In stories, characters usually move between distinct locations, interact with each other and their environment, undergo changes of state, and so on, which is difficult to capture in a single picture. Indeed, picture books tend to have a picture for every 20 to 60 words (cf. Evans & Saint-Aubin, 2005; Verhallen & Bus, 2011). At the same time, a typical visual world display contains fixed objects, which are not designed to match the flow of events. This may cause subjects' attention to the display to decline, perhaps dramatically, after each picture has been identified in relation to the linguistic input. There are at least two studies, however, that suggest otherwise. In the seminal work by Cooper (1974),

participants listened to stories that lasted between 1.5 and 2.5 minutes. Some entities were referenced more than once, but eye movements to the target picture quickly following the referring expression were observed nonetheless. More recently, Richardson and Dale (2005) recorded the eye movements of participants who talked about an episode of The Simpsons or Friends while viewing pictures of these shows' main characters. A 1-minute segment of their speech was later played to listeners who viewed the same set of pictures. It was found that the listener's eye movements closely matched those of the speaker at a delay of approximately 2 sec, and even more closely when a character had just been named. Here, too, repeated reference to the same set of entities did not seem to affect eye movements.

Richardson and Dale (2005) also contributed another relevant observation, which is that the listeners' performance on a recognition test (e.g., *Did the speaker say that Bart electrocuted Homer?*) was correlated with the degree of overlap with the speakers' eye movements. The authors took this to mean that comprehension of spontaneous speech depends on shared visual attention between speaker and listener. Besides that, it may indicate that the most successful comprehenders more quickly moved their eyes to one of the characters after hearing a word that referred to it. In the present study we investigate this link between eye movements and comprehension in more detail.

The Present Study

Six to 11-year-old children listened to a 7-minute story about four animals while viewing pictures of those animals. Although the developments in reading comprehension during this age range are assumed to follow a relatively fixed path, children within an age group may vary considerably. As a consequence, chronological age will only be a crude proxy for an individual's level of development in a given cognitive domain (Bouwmeester, Vermunt, & Sijtsma, 2011). Therefore, we compared children with different levels of performance on a comprehension test, rather than children of different age groups. The test probed children's memory for literal information as well as inferential connections between story events. We used latent class regression analysis (McCutcheon, 1987) to identify classes of children with similar comprehension profiles.

Based on the literature described above, we anticipated three classes: (1) poor performance on both question types, (2) good performance on literal questions but poor performance on inferential questions, and (3) good performance on both question types. The main focus of this research was how these classes would differ from each other with respect to eye movements during listening. Given that resolving discourse anaphora requires inferences, we expect that the class with good inferential comprehension would show a better understanding of anaphoric pronouns, which should manifest itself in more looks toward their pictorial referents. More specifically, we hypothesized the viewing behavior of the poorest comprehenders to be weakly determined by the linguistic input.

That is, compared to the other classes, reference to a story protagonist would less likely be followed by an eye movement to the appropriate picture. In contrast, the class of good literal but poor inferential comprehenders was hypothesized to be more likely to look toward a picture when it was referenced with a name, but not when it was referenced with a pronoun. The class of good literal and inferential comprehenders, finally, should be more likely to look at pictures referenced with nouns and pronouns alike.

Furthermore, we address the question of how children interact with the visual display on a larger time scale. Whereas the pictures will probably attract visual attention in step with the linguistic input at the beginning of the story, this might change later on, varying with comprehension skill in interesting ways. For instance, if children who lack inferential understanding experience cumulative comprehension failure, they will look at the pictures in an increasingly arbitrary manner as the story progresses. A similar pattern, but based on a different mechanism, might be observed for children with good inferential comprehension, in case they require less support from the pictures once they have set up a mental representation into which new information can easily be integrated. Because we know of no prior work addressing the longevity of linguistically-mediated attention to a visual world that allows us to frame explicit hypotheses, this part of the research is exploratory.

Method

Participants

Sixty-nine children from grades 1 to 5 (37 boys, 32 girls, age range 6-11, M = 8.9, SD = 1.5, see Table 1 for an overview of age distributions per grade) in a public primary school participated. Informed consent was obtained from their parents or caretakers. We also obtained consent for using their Cito scores on reading comprehension (measuring children's ability to answer 50 multiple choice questions about passages from a variety of genres) that had been administered independently to children from grade 2 onwards six weeks prior to testing. All participating children had normal or corrected-to-normal vision. Data from five children were not included in the analyses because a language delay was reported after testing. Furthermore, data from one child could not be used because of failure to calibrate the eye-tracker.

Materials

An experimenter-constructed story (see Appendix A) involving four protagonists (a hedgehog, a rabbit, a squirrel, and a mouse) was recorded by a female native speaker of Dutch. The story was 1253 words long and lasted 7 minutes and 44 seconds. It was read in an animated tone, but the same voice was used for all four protagonists when there was dialogue expressed in direct speech. The animals had human-like properties, such as being able to talk or blushing when embarrassed. They were all masculine, so that the

grammatical gender of a pronoun could not be used as a cue. The protagonists were all named and then ascribed some characteristics (e.g., "The squirrel often makes jokes") before the plot unfolded.

The visual display was designed for a 21-inch Tobii 2150 monitor at a resolution of 1600 x 1200 pixels. Line drawings of the four protagonists (see Appendix B) were used, rather than realistic pictures, so that it would be easier to accept their anthropomorphic properties. The line drawings were black-and-white in order to minimize differences in visual salience. Each line drawing was fitted into an area of 400 x 300 pixels, centered within a quadrant of 800 x 600 pixels. Areas of interest (AOIs) were defined as the area of 400 x 300 pixels occupied by the line drawings. The absolute position of the AOIs was fixed, but the distribution of the protagonists over quadrants was randomized between subjects. All protagonists faced outward, so as to avoid seemingly looking at one another.

A 15-item comprehension test was designed to assess comprehension of the narrative. There were seven literal questions that asked for explicitly stated details (e.g., "Which animal overslept?"). Also, there were eight inferential questions that addressed whether participants were able to make certain inferences (e.g., "Why did the hedgehog blush?"). Of these, three questions aimed at a spatial representation of the narrative.[1] Participants were presented with a schematic map of the forest and asked to indicate where certain events had taken place (e.g., "Can you draw a '1' on the place where the rabbit ran into the mouse and the hedgehog?"). For literal questions, one point was given for each correct answer. For inferential questions, one point was given for each correct answer with a valid explanation. That is, if the correct answer was given, children were prompted with the question "Why?" or "How do you know?" A valid explanation consisted of correctly stating the premise for the inference. For the spatial questions, one point was given if the mark was placed on the correct location.

Table 1
Distribution of Latent Class Membership and Age across Grades

		Class membership		Age		
Grade	*n*	n_{Good}	n_{Poor}	Min	Max	Mean
1	11	3	8	6.5	7.9	7.0
2	16	3	13	7.6	9.0	8.3
3	16	4	12	8.2	10.2	9.0
4	7	4	3	9.6	11.1	10.1
5	13	9	4	10.6	11.4	11.2

[1] Although we do not consider these spatial questions as fundamentally different from the other inferential questions, we nevertheless included them separately in the analysis because answering them might require additional skills.

Not all referring expressions in the story were suitable for analysis, for instance because they were genuinely ambiguous, were plural and therefore did not have a unique referent (e.g., 'they'), or were used in conjunction with another referring expression too quickly to allow for a 2-sec time window that was not influenced by other referring expressions (e.g., 'the hedgehog looked at the squirrel'). Also, because they involved more complex narrative constructions, deictic pronouns used in direct speech quotations (e.g., 'I', 'you') were excluded, as were nouns that identified the speaker afterward (e.g. 'said the mouse'). In total, 42 expressions were analyzed (28 names and 14 anaphoric pronouns, underlined in Appendix A). In a pilot study with 14 undergraduate psychology students (ages 18-24, M = 20.2; comprehension test scores 7-15, M = 11.43), these expressions were on average followed by a fixation on the target picture within 2 sec by 76% of the participants (names: M = 82%, SD = 18%, pronouns: M = 65%, SD = 29%). Thus, they served as reliable cues for moving the eyes for adult comprehenders, warranting their use in an experiment with younger participants.

Procedure

Children participated individually in a quiet and normally lit room within the school building. The eye-tracking system used was a Tobii 2150 with a sampling rate of 50 Hz. The session started with the calibration of the eye-tracker, for which participants were seated approximately 60 cm from the screen. Although the setup allowed for relatively free arm and head movement, participants were instructed to sit as still as possible. After calibration, the experimenter explained that they were going to hear an 8-minute story and that they were free to look around on the screen as they listened. They were also told to pay close attention to the narrative, as there would be some questions afterwards. Following the instruction, the visual display appeared. The protagonists were visible for 3 sec before the narrative started.

When the narrative had finished, the comprehension test was administered orally. The verbal responses to the literal and inferential questions were recorded for later reference. For the inferential questions, children were asked to explain their answers. If no answer was given for 5 sec, the experimenter repeated the question. If the child failed to respond or indicated he or she did not know the answer, the experimenter proceeded to the next question. Finally, for the spatial questions, children were provided with a map of the forest. The entire session lasted approximately 20 min.

Results

Comprehension Questions

Scores ranged from 1 to 13 (M = 6.74, SD = 3.45). Cronbach's α for the comprehension test was .71, which is sufficient when the aim is to compare groups. There was a strong correlation (r = .63, p < .001) between performance the comprehension test

and the standardized Cito reading ability scores (M = 9.69, SD = 16.30), establishing the convergent validity of the task.

Latent class analysis. We performed a latent class regression analysis to investigate the number of comprehension classes. The dependent variables (DVs), $\boldsymbol{y}$, are the responses of the children to the questions, collected in a vector. The DVs are dichotomous, because the responses on each question were either correct (1) or incorrect (0). These responses were predicted by a categorical latent variable, x. This latent variable consists of discrete latent classes, which in our case were expected to represent the comprehension levels that drive the children's responses to the questions. The latent classes are assumed to differ in the way the responses are affected by the independent variable question type (Q). This independent variable consists of three levels indicating literal, inferential or spatial questions. Because the DVs are dichotomous, the regression model used has the form of a logit model for the probabilities P(y = 1, xQ). These are the probabilities of providing a correct response to a particular question conditional on a child being in latent class x and responding to question type Q. The logit regression model is defined as follows[2]:

$$\text{Logit P}(y = 1, xQ) = \beta_{0x} + \boldsymbol{\beta}_{1x} Q.$$

The first parameter, β_{0x}, is the intercept for latent class x. The second vector of parameters, $\boldsymbol{\beta}_{1x}$, concerns the regression weights for latent class x for the three question types Q.

We used the program Latent GOLD 4.5 (Vermunt & Magidson, 2008) to perform the analyses of our latent class regression model. We determined the number of latent classes that represented the data best by fitting models with an increasing number of latent classes and interpreted the model that showed the best fit. The log likelihood of a model indicates the fit of the model to the data, with a lower value indicating a better fit (i.e., a smaller difference between the estimated model and the observed data). The number of parameters indicates the parsimony of the model. The balance between fit and parsimony of different models was estimated using the Bayesian Information Criterion (BIC, defined as –2 × log likelihood + number of parameters × ln(N)) and the Akaike Information Criterion 3 (AIC3, defined as –2 × log likelihood + 3 × number of parameters). The model with the lowest value on the information criteria indicates the best balance between fit and parsimony and is the one that should be interpreted. We compared models with one, two, and three latent classes. A model with one class showed the relatively worst fit with the data (L^2 = 765.76, n_{par} = 3, BIC = 517.17, $AIC3$ = 585.76). A model with two classes fitted the data better (L^2 = 684.19, n_{par} = 7, BIC = 452.17, $AIC3$ = 516.19). Adding a third class

[2] The bold regression parameters are vectors, because a parameter had to be estimated for each category of the corresponding variables. Note that the number of free parameters for each categorical independent variable equals the number of categories minus 1.

only marginally improved model fit while increasing the number of parameters (L^2 = 682.65, n_{par} = 11, *BIC* = 467.20, *AIC3* = 526.65). Also, this third class consisted of only three children, making a comparison between classes difficult. We therefore chose to interpret the two-class model, giving up the hypothesized class of good literal and poor inferential comprehension. Table 2 shows the estimated probabilities of giving a correct answer on a particular question type Q within each latent class. The two classes will henceforth be referred to as poor (n = 39, M_{acc} = 4.65, ages 6.5–11.4, M_{age} = 8.7) and good comprehenders (n = 24, M_{acc} = 10.65, ages 6.9–11.4, M_{age} = 9.4). While the good comprehenders were older on average, each grade contributed substantially to both latent classes.

Table 2
Probabilities of Correct Answers on Different Question Types in the One-, Two-, and Three-Class Model

			Question Type β_I			
Model	Latent Class x	n	Literal	Inferential	Spatial	Overall
1-class	1	63	.53	.37	.42	.45
2-class	1	39	.36	.21	.35	.31
	2	24	.82	.63	.54	.70
3-class	1	39	.36	.21	.35	.31
	2	21	.77	.65	.53	.70
	3	3	.99	.53	.31	.70

Eye Movements

Windows of 2000 ms, each consisting of 100 samples of fixation data, following the onset of a referring expression were used for analysis. From the data points thus selected, 20.4% constituted missing values. These could be due to the participant looking away from the screen, blinking, or the eye-tracker failing to record the participant's gaze for technical reasons. Missing values were equally distributed across good and poor comprehenders (19.2% and 21.1%, respectively), but their occurrence increased toward the end of the story (from 16.4% in the first half to 25.1% in the second half), probably reflecting a decrease in attention and less constrained movement on the part of the children.

Multilevel logistic regression. Eye movement data were analyzed in a multilevel logistic regression model (cf. Barr, 2008). This approach allows one to model the effect of a continuous predictor on a categorical outcome, in this case that of time (in increments of 20 ms) on the probability of fixation on the target AOI, while minimizing the need for data aggregation and consequent loss of statistical power. Therefore, the analysis was

performed on the raw eye movement data. On the highest level of the multilevel model fixation on the target picture (y_{ij}) is predicted by time increment *T*:

$$\text{Logit P}(y_{ij} = 1) = \beta_{0i} + \beta_{1i}\, T + e_{ij}.$$

Since *y* is a dichotomous variable (fixation on the target picture as $y = 1$, fixation on any other area as $y = 0$) a logit function is used. β_{0i} is the intercept term that predicts the fixation probability at the onset of a referring expression, β_{1i} is the slope parameter which indicates the change in the fixation probabilities as time *T* increments from 0 to 2 sec (in steps of .02 sec), and *e* is the error term which is assumed to be normally distributed, $e \approx N(0, \sigma^2)$. Both β_{0i} and β_{1i} are in turn defined as variables which form the dependent variables of the regression functions on the lower hierarchical level:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\, W_i + \gamma_{02}\, C + \gamma_{03}\, R_i + \gamma_{04}\, WC_i + \gamma_{05}\, CR_i + u_{0i}.$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}\, W_i + \gamma_{12}\, C + \gamma_{13}\, R_i + \gamma_{14}\, WC_i + \gamma_{15}\, CR_i + u_{1i}.$$

The independent variables for both dependent variables β_{0i} and β_{1i} are window position, *W* (the time point in the story at which the 2-sec window initiated, in minutes, centered[3]), comprehension, *C* (good vs. poor), reference type, *R* (name vs. anaphoric pronoun), and the interactions Time x Comprehension and Comprehension x Reference Type.[4] Both u_0 and u_1 are error terms which are assumed to be normally distributed, $u \approx N(0, \sigma^2)$ and are allowed to correlate with each other. The parameter estimates are given in Table 3 and displayed graphically in Figure 1. For the onset β_{0i}, there was a main effect of window position, $\gamma_{01} = -.120$, $SE = .006$, $Wald(1) = 421.852$, $p < .001$. The further the story progressed, the lower the probability that a child fixated the target picture at word onset. There was a significant main effect of comprehension, $\gamma_{02} = .141$, $SE = .015$, $Wald(1) = 93.498$, $p < .001$, indicating that good comprehenders were more likely to make anticipatory on-target fixations than poor comprehenders. There was a significant main effect of reference type, $\gamma_{03} = -.238$, $SE = .013$, $Wald(1) = 351.814$, $p < .001$, indicating a lower probability of anticipatory on-target fixations for names than for pronouns. There was a significant interaction effect of Window Position x Comprehension, $\gamma_{04} = -.043$, $SE = .006$, $Wald(1) = 53.741$, $p < .001$. As the story progressed, the likelihood of anticipatory on-target fixations decreased more strongly for good comprehenders than for poor

[3] Window position was centered so that the parameter estimates reflect the fixation probabilities as children were in the middle of the story. This was done to avoid a bias toward viewing behavior at the beginning of the story in the interpretation of the results.

[4] Because we did not have hypotheses about the interaction Window Position x Reference Type, or about the three-way interaction Window Position x Comprehension x Reference Type, we left these out of the model to restrict the number of parameters.

comprehenders. Finally, there was a significant interaction effect of Comprehension x Reference Type, γ_{05} = -.070, *SE* = .013, *Wald*(1) = 30.730, $p < .001$, indicating that the difference between names and pronouns was largest for good comprehenders.

For the slope β_{1i}, the intercept was significant, γ_{10} = .178, *SE* = .012, *Wald*(1) = 218.254, $p < .001$, meaning that the probability that a child fixated the target picture increased during the 2 sec after the onset of a referring expression. There was a main effect of window position, γ_{11} = -.031, *SE* = .005, *Wald*(1) = 35.362, p < .001, indicating that the increase of the probability of on-target fixation during the 2 sec after word onset was attenuated as the story progressed. The main effect of comprehension was not significant, γ_{12} = -.005, *SE* = .012, *Wald*(1) = .202, p = .65, suggesting that the slopes were similar for good and poor comprehenders. There was a significant main effect of reference type, γ_{13} = .159, *SE* = .011, *Wald*(1) = 213.238, $p < .001$. The probability that a child fixated the target picture increased more strongly after a name than after a pronoun. There was a significant interaction effect of Window Position x Comprehension, γ_{14} = -.025, *SE* = .005, *Wald*(1) = 22.433, $p < .001$, indicating that as the story progressed, the increase in fixation probability following a referring expression was attenuated more strongly for good comprehenders than for poor comprehenders. Finally, the interaction effect of Comprehension x Reference Type was not significant, γ_{15} = .006, *SE* = .011, *Wald*(1) = .346, $p = .56$, suggesting that the difference between the slopes for names and pronouns was similar across good and poor comprehenders.

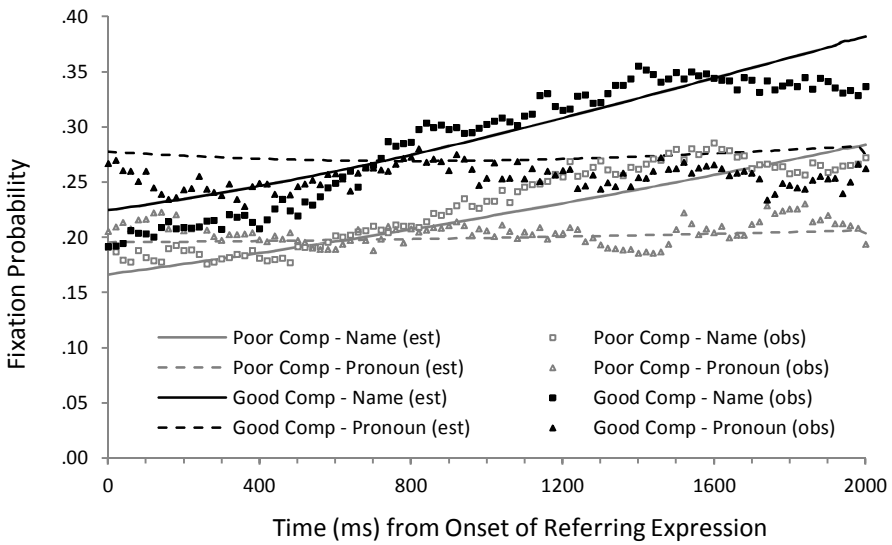


Figure 1. Estimated (est) and observed (obs) average fixation probabilities for the target picture between 0 and 2000 ms after onset of a referring expression.

Table 3
Estimates of the Fixed and Random Effects of the Multilevel Logistic Regression Model of Fixation Probabilities for Target Pictures

Term	Predictor	Estimate	*SE*	Wald	df	*p*
β_0 Onset	γ_{00} Intercept	-1.377	.015	8409.569	1	<.001
	γ_{01} Window Position	-.120	.006	421.852	1	<.001
	γ_{02} Comprehension	.141	.015	93.498	1	<.001
	γ_{03} Reference Type	-.238	.013	351.814	1	<.001
	γ_{04} Window Position x Comprehension	-.043	.006	53.741	1	<.001
	γ_{06} Comprehension x Reference Type	-.070	.013	30.730	1	<.001
	u_{0i} Random Effect	.225	.005	2081.046	1	<.001
β_1 Slope	γ_{10} Intercept	.178	.012	218.254	1	<.000
	γ_{11} Window Position	-.031	.005	35.362	1	<.001
	γ_{12} Comprehension	.005	.012	.202	1	.650
	γ_{13} Reference Type	.159	.011	213.238	1	<.001
	γ_{14} Window Position x Comprehension	-.025	.005	22.433	1	<.001
	γ_{15} Comprehension x Reference Type	.006	.011	.345	1	.560
	u_{1i} Random Effect	.162	.004	1374.202	1	<.001

Note. Comprehension (contrast coding): good (1) versus poor (-1); Reference (contrast coding): name (1) versus pronoun (-1).

Baseline comparisons. Between 0 and 2000 ms after the onset of a referring expression, good comprehenders were more likely to fixate on the target picture than poor comprehenders. But does this difference reflect linguistically-mediated visual attention, or is it a consequence of the fact that good comprehenders were more likely to look at *any* of the pictures throughout the story? To find out, we compared our results against a baseline of children's viewing tendencies. This baseline was obtained by matching the eye movement data to a new scoring scheme. Instead of analyzing fixations on the designated target picture, we analyzed the fixations on a picture that was randomly drawn from the set of four pictures for each of the 42 time windows. For instance, for the 2 sec-window following *hedgehog*, a fixation on the rabbit was scored as 1, and on any other area as 0. A more detailed account of the procedure and the statistical tests is provided in Appendix C. In short, the baseline comparisons show that all children were significantly more likely to fixate target pictures than randomly selected pictures, and that only for target pictures clear differences between good and poor comprehenders were observed. Thus, the differences between good and poor comprehenders' eye movements can be attributed to their processing of the concurrent linguistic input. We will address these differences in detail in the discussion.

Age and comprehension skill. We used latent class membership, rather than age, as an approximation of children's comprehension skill. This approach was supported by the observation that the latent classes run through all grades (see Table 2). Yet, it could be argued that some of the differences in online viewing behavior are driven by developmental differences beyond those measured by the comprehension test, such as better oculomotor control (Yang, Bucci, & Kapoula, 2002) or attentional persistence. To address this issue, we performed an additional analysis, in which we entered age (continuous, in months) as a covariate in addition to the comprehension classes. A description of the model and the statistical tests is given in Appendix D. The most important result was that comprehension skill still explained unique variance in the onset term, as did the interaction between comprehension and reference type. This suggests that the effects of comprehension cannot be explained in terms of age alone. The regression weight of comprehension skill even increased from .141 to .183, and that of the interaction from -.070 to -.085. Older age was associated with a lower probability of fixating the target picture at word onset, but this effect was qualified by an interaction with window position, such that older children showed a smaller decline as the story progressed. Finally, older age was associated with steeper slopes, indicating that older children were more likely to shift their gaze toward the target picture in the 2000 ms following a referring expression than younger children.

Discussion

The aim of this study was to increase our understanding of children's online processing of a narrative in relation to comprehension outcomes. We used a 7-minute story, allowing us to adhere closely to the task demands of story comprehension, which involve dealing with a continuous stream of information, keeping track of multiple characters, and establishing coherence in a multitude of other dimensions, such as causality, time, and space (Zwaan & Radvansky, 1998). As such, the task provides a potentially more valid test of children's ability to keep track of discourse entities than previous work. Furthermore, the rich and complex nature of the story allowed us to assess in detail the children's memory representation of the text, rather than use some external measure of reading comprehension.

Applying latent class analysis to the children's performance on a set of literal and inferential questions, a solution with two classes, defined by good and poor overall comprehension of the story, provided the best fit with the data. The absence of a class that performed well on literal questions while showing poor inferential comprehension was somewhat unexpected, given that such a pattern has been observed in other studies (e.g., Cain & Oakhill, 1999; Oakhill, 1982; Omanson et al., 1978; Zabrucky & Ratner, 1986; but see Oakhill, 1984 for outcomes similar to ours). The three-class solution did add just such a class (see Table 1), but described the data only slightly better while being less

parsimonious, and therefore less likely to generalize beyond the sample. This discrepancy might be due to the relative difficulty of the task. Answering the literal questions required retention of story details for up to seven minutes, which is markedly longer than in most text comprehension research with young children. For this reason, perhaps, the task did not discriminate among levels of suboptimal comprehension as well as some other tasks. Still, it should be noted that there were children in all grades who performed well on the comprehension test, suggesting that although the story was quite sophisticated, the task was not inappropriately difficult.

Despite these concerns, it was possible to meaningfully compare the two classes of comprehenders. These should differ from each other in two ways. First, the poor comprehenders were expected to experience a general weakness in directing their visual attention in step with the linguistic input, such that they should be less likely to look at the target picture than the good comprehenders, both after context-independent reference (i.e., proper names) and context-dependent reference (i.e., anaphoric pronouns). Second, the generation of appropriate inferences and the use of discourse-level cues to referential solution were expected to constitute an additional source of difficulty, so the difference with good comprehenders should be greatest for anaphoric pronouns.

The eye movement data support both a general and a discourse-level explanation of differences in linking the linguistic input to the visual world. As Figure 1 shows, under the scope of names and anaphoric pronouns alike, good comprehenders were more likely to look at the target picture than poor comprehenders. This was not because they were more likely to look at *any* of the pictures, as the baseline comparisons showed. For the slope term of the model, there were no significant differences between good and poor comprehenders, suggesting that both classes similarly adjusted their gaze preference in response to hearing a referring expression. At first, this seems to run counter to the discourse-level explanation, but it should also be noted that in both classes anaphoric pronouns led to a significantly smaller increase in fixation probability (i.e., a shallower, virtually flat slope) than proper names. This is consistent with the notion that anaphoric pronouns are generally used to refer to a discourse entity that is already highly accessible to both the speaker and the listener and has a relatively high likelihood of being mentioned again, while a name is used to bring a currently inaccessible entity into focus (Clark & Wilkes-Gibbs, 1986; Givón, 1992). In the latter case, it is inherently more difficult to predict which entity is coming up. The crucial comparison, then, lies in the probabilities of fixating the target picture at word onset. Rather than treat these early fixations as artifacts of the pictures being visible before the critical word, we see them as reflecting listeners' genuine expectations about what a speaker might mention (Barr, Gann, & Pierce, 2011). Indeed, good comprehenders were more likely to make anticipatory fixations on the target picture than poor comprehenders, and even more so when the picture was referenced by an anaphoric pronoun.

Precisely how can we account for the different levels of anticipation for good and poor comprehenders? We propose two possible mechanisms. First, according to the expectancy hypothesis (Arnold, 2001), listeners constantly monitor a variety of cues to anticipate whether currently accessible information will remain important to the speaker. Among such cues are order-of-mention, grammatical subjecthood, and even subtle hesitations (e.g., 'uh'; Arnold, Tanenhaus, Altmann, & Fagnano, 2004). A greater ability to notice and apply these forward-looking cues is a plausible explanation for why children with good comprehension anticipated the referents of pronouns to a greater extent than children with poor comprehension. Consequently, making more accurate judgments about who did what to whom throughout the story partly explains their superior performance on the comprehension test.

Second, inferencing ability may play a role. Distributional cues help shape predictions about upcoming information, but they are not sufficient by themselves. As in the relatively simple example of *Jane needed Susan's pencil. She gave it to her* discussed in the introduction, *some* knowledge of how events in the world are related is often necessary to determine who did what to whom. So, alternatively, poor comprehenders might not make the necessary inferences during listening to be able to determine a pronoun's referent. This possibility is supported by their low (.21) performance on inferential questions compared to good comprehenders (.64), and by previous research that points to weak inferential skills as a cause of poor anaphoric comprehension (e.g., Oakhill & Yuill, 1986). However, this type of inferential processing, which requires a search in memory for a suitable antecedent, may have a distinct be temporal profile, in that it should pull eye movements *after* hearing the pronoun. Given that the slopes of the model did not differ between good and poor comprehenders, our findings do not provide direct support for differential ability in this regard. It is possible, though, that certain inferences constrain expectations regarding which discourse entities were most likely to be mentioned next. There is evidence that the accessibility of entities during comprehension is mediated by spatial co-presence in the described situation (Glenberg, Meyer, & Lindem, 1987; Nieuwland, Otten, & Van Berkum, 2007). By the same token, when the story switches to the journey of the hedgehog and the mouse through one part of the forest, it is not likely that the pronoun *he* would refer to the squirrel, who is left behind in another. So, an accurate spatial model of the narrated situation, as probed by some of the items in the comprehension test, should facilitate referential solution.

Given the data, a combination of the two anticipatory mechanisms is possible too. As Ackerman (1986) points out, explanations in terms of familiarity with the textual devices that establish coherence on the one hand and inferencing ability on the other are not mutually exclusive and may even reinforce one another. For instance, when forward-looking distributional cues are not noticed, the load on inferential processing to determine a pronoun's referent increases. A rather different explanation pertains to working

memory. The likelihood that one keeps in mind which entities are likely to remain important for the foreseeable future of the story while processing other information might be expressed as a function of working memory capacity, which we did not control for Although this may have been informative, it would not discount anticipation as a mechanism behind the observed gazing behavior; it moves the emphasis from possessing the skill to generate expectations based on the input to bringing these expectations to bear on referential choice. In either case, further research is needed to tease out precisely which information in the discourse good comprehenders took advantage of and poor comprehenders failed to take into account for their predictions of the narrative focus. The present results provide a first hint that these differences are key to understanding the processes that constitute successful comprehension.

We also explored the effect of the position of the referring expression within the story on the probability of looks toward the target pictures. Did children's viewing behavior change as the story progressed? It did, in a couple of ways. First, over time, children became less likely to fixate the target picture at word onset. This was especially true for good comprehenders. Second, the slopes also decreased somewhat as the story progressed, again mainly so for good comprehenders. So, good and poor comprehenders became more similar in their viewing behavior the further they got into the task. This pattern rules out an explanation of the two classes emerging due to a differential role of fatigue or loss of attention. If anything, fatigue made the classes behave more alike. A possibility is that good comprehenders understood less as the story progressed, but their accuracy on questions targeting the first and second half of the story was comparable (.80 and .67, respectively[5]). To account for this observation, we speculate that there exist two mechanisms that drive endogenous eye movements (which reflect task-driven orienting, as opposed to exogenous eye movements, which reflect stimulus-driven orienting; Jonides, 1981) during listening, and that these were active during different phases of the task.

At first, the pictures may be helpful in providing a model on which the mental representations of the characters are based. When the speaker refers to one of the pictures, this serves as a strong incentive to make an eye movement to that picture. It is during this phase that listeners engage in a *motivated* search for the appropriate picture and that eye movements are most susceptible to underlying differences in referential processing ability. This would explain why the difference between good and poor comprehenders was greatest during the beginning of the story. But whereas pictures function as a scaffold for comprehension initially, their supporting role becomes less obvious later on, given that they do not provide any additional information about the scenes and at times even

[5] These numbers are computed over a subset of 11 questions for which the answers could only be found in one specific part of the text. Therefore, they do not add up to the numbers in Table 1.

contradict information in the story (e.g., when the rabbit is described as running, or the squirrel as lying down). Consequently, listeners will no longer look at the pictures in the expectation that they augment the linguistic input.

From then onwards, eye movements may be more about using spatial structure than getting visual data, as Richardson and Dale (2005) propose. That is, the display serves as an external memory store, in which each of the four protagonists is associated with an oculomotor coordinate (which was encoded during the first phase). Hearing 'the rabbit', for example, automatically activates a representation that includes the spatial coordinates of the rabbit in its expected location, causing the eyes to move to that location, regardless of whether the visual information is relevant at that time (see Altmann, 2004, for a similar discussion of eye movements to a blank screen). Apparently, this second mechanism is weaker, at least in children, but their viewing behavior during the latter part of the story is not completely arbitrary either, and still reflects linguistic processing to a certain extent. Of course, further research will be needed to evaluate these causal interpretations. Overall, it seems that during long exposure to simple pictures, the relationship between eye movements and ongoing comprehension processes becomes gradually less clear. These findings provide an interesting extension to earlier work by Cooper (1974) and Richardson and Dale (2005) and suggest that for maximizing the effects of underlying processing differences, a particular visual array should be used for a few minutes at most.

To conclude, the present work adds to our understanding of children's text comprehension in several ways. First, it shows that performance on an offline comprehension task is associated with specific patterns of online referential processing. Quite surprisingly, good and poor comprehenders did not differ significantly in the adjustment of their viewing preferences *after* hearing a referring expression. However, good comprehenders were more likely to look at the picture of the protagonist in narrative focus *before* it was mentioned, especially when it was referenced with an anaphoric pronoun. Conversely, poor comprehenders seemed to be less 'tuned in' at anticipating whether or not a presently accessible protagonist would continue to be important in the story. As such, the present results put emphasis on the role of prediction during discourse comprehension, aligning with findings from adults' language comprehension (e.g., Altmann & Kamide, 2007; Kamide, Altmann, & Haywood, 2004). Second, it shows that latent classes based on a comprehension test are related to different patterns of online viewing behavior, independently of age. While an individual's age was associated with certain characteristics of gazing behavior, such as the probability of shifting gaze toward the target picture after hearing a referring expression, it was not the case that older children necessarily showed more anticipatory viewing behavior. Third, and finally, the present work extends findings from earlier eye-tracking research in referential processing by showing that effects of comprehension skill on referential processing can be observed

in a naturalistic discourse task. So, directing anticipatory eye movements to a visual world does not seem to be restricted to short descriptions of a scene, but something that proficient comprehenders do continuously while listening to a story.

Appendix A
The Story (Translated from Dutch)

Eye movements concurrent with underlined words were analyzed.

This story is about the hedgehog, the rabbit, the squirrel and the mouse. They are good friends of each other. The hedgehog is a little quiet and shy. The rabbit is a fast runner. The squirrel makes a lot of jokes. The mouse is always looking for adventure. They are always into something together. They often play near the lake or the giant rock.

One day the mouse was walking through the forest. He was a little bored. He decided to go and visit the hedgehog. That was just a short walk. When he had arrived at the hedgehog's place, he knocked. After a long time, the hedgehog opened. The mouse laughed. The hedgehog looked rather sleepy. "I overslept, I guess", the hedgehog said. "Yes. Again", the mouse said. "Will you join me for a walk?" "Okay", the hedgehog said. "Let me brush my teeth first."

Half an hour later, the hedgehog and the mouse were walking on the big forest road. They discussed their plans for the day. "I wonder where the other two are", the hedgehog said. "The rabbit and the squirrel?" the mouse asked. "Yes", the hedgehog replied. "They're bound to be at the lake", the mouse said. The hedgehog proposed going there too. The mouse thought it was a good idea. But it would be a long journey if they were to take the big forest road. The mouse said he knew a shortcut. They needed to turn right when they arrived at the giant rock. "But won't we get lost?" the hedgehog asked. "No one has ever got lost with me", the mouse said. "Alright then," the hedgehog said. When they arrived at the giant rock, they took a right turn into the bushes.

Meanwhile at the lake, the rabbit and the squirrel were sitting on a log. The squirrel wanted to do a running contest with the rabbit. "I'm sure that from here I can run to the giant rock and back faster than you", he said. "Well, let's see", the rabbit said. "Okay", the squirrel said. "I'll count to three and at three we run". They both got on their marks. The squirrel started counting: "One… two… three!" The rabbit dashed away with great speed. But what did the squirrel do? He just stayed there. The rabbit didn't notice anything and rushed on. The squirrel lay down on the log in the sun. He thought it was a good joke and knew what he'd say when his friend would come back.

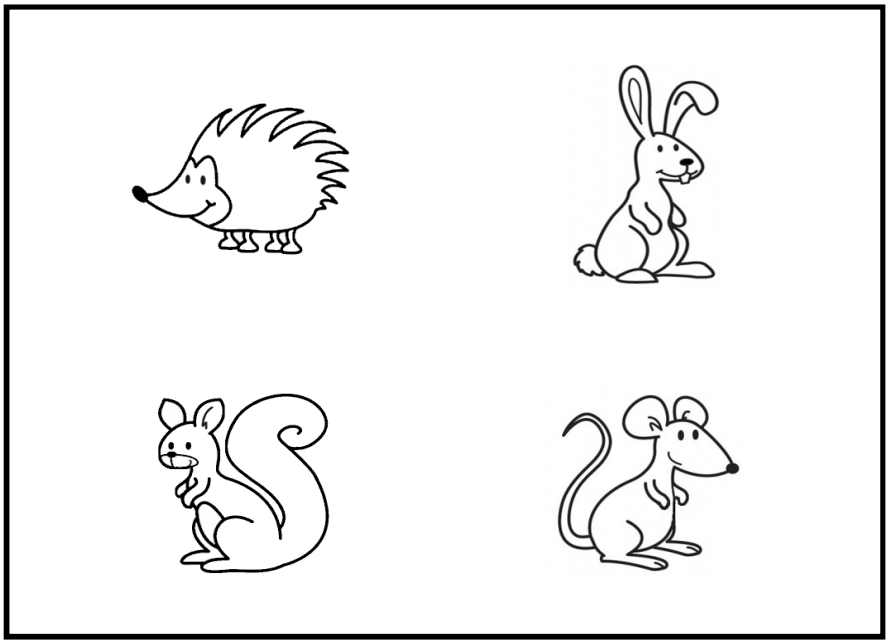
Elsewhere in the forest, the mouse and the hedgehog were still on their way. They had strayed from the forest road quite a bit. "I'm sure I passed this spruce last time I went to the lake", the mouse said. "But all the trees here look alike", the hedgehog said. "I think we're lost." "Come on", the mouse said bravely, "It can't be far". "Why didn't we just take the forest road?" the hedgehog whined. "Maybe we'll never find the way back". "Sssh", the mouse said. "I hear something." They stopped to listen. It sounded like footsteps of an animal fast approaching. The hedgehog quickly jumped into the bushes to hide. The mouse

followed suit. They held their breath as the steps came closer. Then they saw the rabbit running past. The hedgehog came out of the bushes immediately. “Rabbit, rabbit!” he called. The rabbit stopped. “Why are you in such a hurry?” the hedgehog asked. “Hello hedgehog”, the rabbit said. “Squirrel and I are doing a contest. But I can’t see him anywhere. I must be way ahead.” “Hello rabbit”, said the mouse, who had left the bushes too. “We were looking for you and squirrel. We thought you’d be at the lake.” “Yes”, the rabbit said. “We’re doing a running contest from the lake to the giant rock and back. So if you want to play with us, you’d best walk to the lake.” “So, to get to the lake, this would be the shortest way?” the mouse asked. “Yes, you’re almost there. Why?” the rabbit asked. “Oh, just because”, the mouse said. The hedgehog blushed. “Well, I’m off again. Send my regards to the squirrel when you guys see him”, the rabbit said.

On the log near the lake the squirrel lay waiting calmly in the sun. But he started to get a little worried. He got up and started hopping up and down the log. The rabbit should have been back by now, he told himself. Might something have happened on the way? Maybe it wasn’t that good a joke after all. He wondered whether he should go looking for the rabbit. At that moment, the hedgehog and the mouse emerged from the bushes. The squirrel got up immediately. “Hey, you!” he shouted. Quickly the mouse whispered to the hedgehog: “I’m sure that squirrel has played a trick on rabbit. Let’s get back at him.” The hedgehog nodded silently. “Hello squirrel!” the mouse said in a loud voice. “Have you seen rabbit?” the squirrel asked. “Rabbit?” the mouse replied. “No, we thought he’d be here with you. Where did he go?” “Err…” the squirrel stammered. “Err… He just left for the giant rock. To collect something.” “Anyway, we didn’t see him”, the mouse said. “But he left through the same bushes you just came from. He wouldn’t be lost, would he?” the squirrel asked. “Well, I’m sorry”, the mouse said. “Perhaps he got stuck in a blackberry bush”. “Oh no!” the squirrel cried. “But was he in a hurry? Otherwise he’d be more careful, right?” the mouse said. “Err...” the squirrel hesitated. “Yes, he was in a hurry, I guess.” And then he decided to confess the prank he had pulled. He was very sorry and started to weep quietly.

At that moment the rabbit came dashing out of the bushes. He stopped and looked very surprised when he saw the squirrel. “Hey! How come you’re here already? And hey, why are you crying?” “Oh, I’m so glad to see you, rabbit!” the squirrel said. “I pulled your leg. I stayed behind at the log to lie in the sun. But then I got afraid you had got lost.” The rabbit thought for a moment. The he replied: “Well, it just took a little longer because I met mouse and squirrel on the way. The squirrel looked at the mouse and the hedgehog. “Why didn’t you tell?” “We thought you were pulling someone’s leg again”, the hedgehog said. “And we wanted to get back at you. But we didn’t mean to upset you.” “You got just what you deserved”, the mouse said. “And you’ve lost the running contest,” the rabbit laughed. “I won’t do it again”, the squirrel promised, and dried his tears. “Shall we play hide and seek now?” And so the friends were together again, a whole day of fun ahead of them. They no longer tried to fool each other.

Appendix B
The Visual World Display



Appendix C
Description of Baseline Analyses

To rule out the possibility that good comprehenders simply looked toward the pictures on the screen more than poor comprehenders, and as a result were found to also look at a given picture more when it happened to be referenced in the story, we estimated a baseline of visual attention. To do so, we created a new dependent variable by randomly assigning (via a function in Excel) one of the four pictures to the time windows following a referring expression and computing the probability that this picture was fixated.

Table C1 gives a sample dataset illustrating this approach. When the observed fixation was on the picture designated as the target picture in the story, *Fix=Target* received the value 1. When the observed fixation was on the picture that happened to be assigned to that trial, the value *Fix=Random* received the value 1. The values of the newly calculated variable were added as independent observations to each participant's eye movement data. If the viewing tendencies of good and poor comprehenders differ from each other as a function of linguistic processing, we should find differences in the probability of fixating target pictures over and above this baseline. More specifically, the fixation probabilities for the target pictures should be greater than for the randomly selected pictures, and show a greater contrast between good and poor comprehenders.

Table C1
Example Dataset with Fixations on Target and Random Pictures

Subject	Window	Cycle	Target Picture	Random Picture	Observed Fixation	Fix=Target	Fix=Random
1	1	1	Mouse	Squirrel	Mouse	1	0
1	1	2	Mouse	Squirrel	Squirrel	0	1
1	2	1	Rabbit	Mouse	Hedgehog	0	0
1	2	2	Rabbit	Mouse	Hedgehog	0	0
2	1	1	Mouse	Hedgehog	Mouse	1	0
2	1	2	Mouse	Hedgehog	Mouse	1	0
2	2	1	Rabbit	Rabbit	Rabbit	1	1
2	2	2	Rabbit	Rabbit	Mouse	0	0

To assess the effect of picture relevance, we added the predictor picture, *P* (target vs. random) and the interaction Picture x Comprehension to the multilevel logistic regression model described in the Results section of the main text. This yielded the following regression equation:

$$\beta_{0i} = \gamma_{0} + \gamma_{1} W_i + \gamma_{2} P_i + \gamma_{3} C + \gamma_{4} R_i + \gamma_{5} WC_i + \gamma_{6} PC_i + \gamma_{7} CR_i + u_{0i}.$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11} W_i + \gamma_{12} P_i + \gamma_{13} C + \gamma_{14} R_i + \gamma_{15} WC_i + \gamma_{16} PC_i + \gamma_{17} CR_i + u_{1i}.$$

The parameter estimates are given in Table C2. Here we report the added effects of picture type compared to the original model. For the onset β_{0i}, there was a significant effect of picture type, $\gamma_{02} = .161$, $SE = .008$, $Wald(1) = 374.768$, $p < .001$, indicating that children were more likely to fixate the target picture than a randomly selected picture. This effect was qualified by a significant interaction of Picture x Comprehension, $\gamma_{06} = .028$, $SE = .008$, $Wald(1) = 10.932$, $p < .001$, indicating that the advantage for target pictures was largest for good comprehenders. For the slope β_{1i}, there was a main effect of picture, $\gamma_{12} = .120$, $SE = .007$, $Wald(1) = 267.863$, $p < .001$, indicating that the slopes were steeper for target pictures than for random pictures. Thus, in the 2 sec following word onset, the probability of fixating the target picture increased more strongly than the probability of fixating a random picture. This effect was qualified by a significant interaction of Picture x Comprehension, $\gamma_{16} = .019$, $SE = .007$, $Wald(1) = 6.980$, $p = .006$, suggesting that the slope advantage for target pictures was largest for good comprehenders. This can clearly be seen in Figure C1, which shows that the fixation probability for randomly selected pictures did not increase after onset of a referring expression.

In sum, good comprehenders were more likely than poor comprehenders to fixate the picture that was referenced in the story, even when compared to a baseline of their own viewing tendencies. These analyses support the hypothesis that linguistic processing accounts for differences in fixation probability for the target picture over and above other differences that may drive good and poor comprehenders' eye movements.

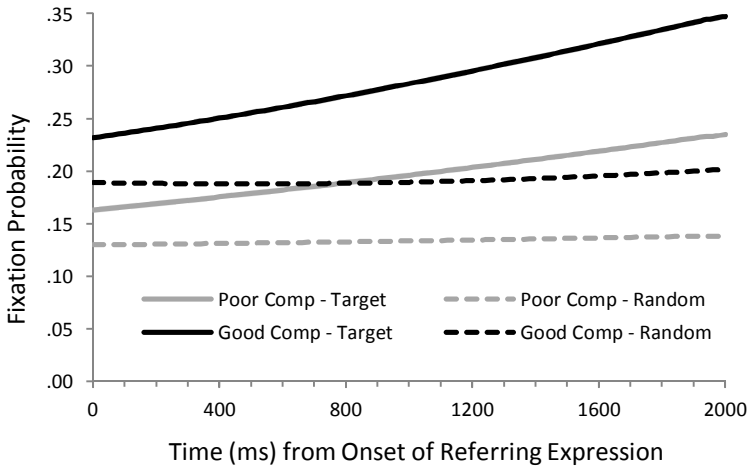


Figure C1. Estimated fixation probabilities for the target picture and a randomly selected picture between 0 and 2000 ms after onset of a referring expression.

Table C2

Estimates of the Fixed and Random Effects of the Multilevel Logistic Regression Model of Fixation Probabilities for Target and Random Pictures

Term	Predictor	Estimate	*SE*	Wald *Z*	df	*p*
β_0 Onset	γ_{00} Intercept	-1.600	.010	25593.009	1	<.0001
	γ_{01} Window Position	-.036	.004	67.019	1	<.0001
	γ_{02} Picture	.161	.008	374.768	1	<.0001
	γ_{03} Comprehension	.152	.010	239.195	1	<.0001
	γ_{04} Reference Type	-.146	.010	236.285	1	<.0001
	γ_{05} Window Position x Comprehension	.017	.004	15.102	1	<.001
	γ_{06} Picture x Comprehension	.028	.008	10.932	1	<.001
	γ_{07} Comprehension x Reference Type	-.032	.010	11.034	1	<.001
	u_{0i} Random Effect	.193	.004	2549.825	1	<.0001
β_1 Slope	γ_{10} Intercept	.123	.009	183.099	1	<.0001
	γ_{11} Window Position	-.018	.004	22.557	1	<.0001
	γ_{12} Picture	.116	.007	267.863	1	<.0001
	γ_{13} Comprehension	.019	.009	4.584	1	.032
	γ_{14} Reference Type	.100	.008	153.581	1	<.0001
	γ_{15} Window Position x Comprehension	-.015	.004	14.990	1	<.001
	γ_{16} Picture x Comprehension	.019	.007	6.980	1	.008
	γ_{17} Comprehension x Reference Type	.011	.008	1.876	1	.170
	u_{1i} Random Effect	.075	.002	1932.727	1	< .0001

Note. Picture (contrast coding): target (1) versus random (-1); Comprehension (contrast coding): good (1) versus poor (-1); Reference (contrast coding): name (1) versus pronoun (-1).

Appendix D
Analyses Using Age

We performed an analysis in which chronological age was added as a covariate in addition to comprehension class. The rationale was that if uncontrolled age differences between the comprehension classes accounted for the differences in the viewing behavior, there should be no unique variance to be explained by the comprehension classes. We included the predictor age, *A*, and its interaction with window position in the multilevel logistic regression model reported in the main text. This yielded the following regression equation:

$$\beta_{0i} = \gamma_{00} + \gamma_{01} W_i + \gamma_{02} C + \gamma_{03} A + \gamma_{04} R_i + \gamma_{05} WC_i + \gamma_{06} WA_i + \gamma_{07} CR_i + u_{0i}.$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11} W_i + \gamma_{12} C + \gamma_{13} A + \gamma_{14} R_i + \gamma_{15} WC_i + \gamma_{16} WA_i + \gamma_{17} CR_i + u_{1i}.$$

The parameter estimates are given in Table D1. We report the meaningful deviations from the standard model. For the onset term, there was a significant effect of age, γ_{03} = -.072, *SE* = .009, *Wald*(1) = 72.686, $p < .001$. The negative coefficient suggests that older children showed a lower probability of fixating the target picture at word onset than younger children. However, there was a significant interaction of Window Position x Age, γ_{06} = .019, *SE* = .006, *Wald*(1) = 11.096, $p < .001$, indicating that older children showed a smaller decline in the probability of fixating a target picture at word onset than younger children. For the slope term of the model, there was a significant effect of age, γ_{13} = .060, *SE* = .007, *Wald*(1) = 71.613, $p < .001$, suggesting that older children were more likely to shift their gaze toward the target picture in the 2000 ms following a referring expression than younger children. The effect of window position was no longer significant, γ_{11} = .061, *SE* = .040, *Wald*(1) = 2.310, $p = .130$, as was the interaction of Window Position x Comprehension, γ_{15} = .007, *SE* = .007, *Wald*(1) = 1.035, $p = .310$. Also, the newly added interaction Window Position x Age was not significant, γ_{16} = -.004, *SE* = .005, *Wald*(1) = .785 $p < .380$. This suggests that factoring in age made that there was no unique variance in the increase of looks to the target picture to be explained by the position of a referring expression within the story. Finally, there was a significant interaction for Comprehension x Reference Type, γ_{17} = .055, *SE* = .012, *Wald*(1) = 21.669, $p < .001$, indicating that for good comprehenders, the difference in slopes for names and anaphoric pronouns was greater than for poor comprehenders.

The main conclusion from these analyses is that only comprehension seems to be associated with different levels of anticipation for names and anaphoric pronouns. Also, when age is included as an additional predictor besides comprehension, the latter still accounts for unique variance. Age does make an independent contribution in explaining

some aspects of viewing behavior, but it does not seem to be strongly related to the effect of theoretical interest.

Table D1
Estimates of the Fixed and Random Effects of the Multilevel Logistic Regression Model of Fixation Probabilities for Target Pictures, Controlling for Age

Term	Predictor	Estimate	*SE*	Wald	df	*p*
β_0 Onset	γ_{00} Intercept	-.700	.077	83.641	1	<.0001
	γ_{01} Window Position	-.187	.049	14.445	1	.0001
	γ_{02} Comprehension	.183	.013	197.802	1	<.0001
	γ_{03}Age	-.072	.009	72.686	1	<.0001
	γ_{04} Reference Type	-.315	.014	511.784	1	<.0001
	γ_{05} Window Position x Comprehension	.062	.009	51.010	1	<.0001
	γ_{06} Window Position x Age	.019	.006	11.096	1	.0009
	γ_{07} Comprehension x Reference Type	-.085	.014	36.997	1	<.0001
	u_{0i} Random Effect	.017	.001	1314.021	1	<.0001
β_1 Slope	γ_{10} Intercept	-.314	.064	24.332	1	<.0001
	γ_{11} Window Position	.061	.040	2.310	1	.130
	γ_{12} Comprehension	-.009	.011	.604	1	.440
	γ_{13}Age	.060	.007	71.631	1	<.0001
	γ_{14} Reference Type	.250	.012	454.645	1	<.0001
	γ_{15} Window Position x Comprehension	.007	.007	1.035	1	.310
	γ_{16} Window Position x Age	-.004	.005	.785	1	.380
	γ_{17} Comprehension x Reference Type	.055	.012	21.669	1	<.0001
	u_{1i} Random Effect	.007	<.001	929.564	1	<.0001

Note. Picture (contrast coding): target (1) versus random (-1); Comprehension (contrast coding): good (1) versus poor (-1); Reference (contrast coding): name (1) versus pronoun (-1).

Chapter 4

The Role of Grounded Event Representations in Discourse Comprehension*

* This chapter has been submitted for publication as Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (submitted). The role of grounded event representations in discourse comprehension.

Abstract

What is the role of grounded event representations in discourse comprehension? While many studies have documented mental simulations at the sentence level, it remains unclear to what extent they support the formation of a coherent situation model. One might even argue that activating grounded representations can interfere with building coherence across sentences, because both require attentional resources. We investigated this issue in two eye-tracking experiments. In Experiment 1, participants listened to short stories while concurrently viewing visual scenes. During target sentences, one picture matched the agent, but not the action, while another picture matched the action, but not the agent. The order in which these pictures were referenced was varied. Participants showed a strong preference for inspecting the agent across word orders. Experiment 2 replicated these findings with a slightly different secondary task. These results suggest that language users prioritize constructing a referentially coherent situation model over forming grounded representations of the actions that are described in them.

John and Bill were walking along a canal when they heard a scream. As they looked around, they saw a child nearly drowning. John did not hesitate and dove into the water. He reached the child with a few strokes and managed to pull it onto the bank.

How is a passage such as the above comprehended? A widely held view is that comprehenders construct a situation model, that is, a mental representation of the events described by the text, rather than of the text itself (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). When a text contains multiple events, these are held together by *coherence relations*. According to the event-indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995), there are at least five dimensions on which coherence is monitored: time, space, entities, causation, and intentionality. In the example above, the first clause introduces two protagonists, a spatial setting, and a temporally protracted event ('walking') that serves as background for a punctuated focal event ('heard a scream'). The second sentence refers back to the protagonists with a pronoun, thus creating coherence on the referential level. Given that there is no explicit change of spatial location, the reader assumes that the next event takes place in the same location (although the protagonists' attention, and with that the reader's attention, is directed to a specific part of the scene). Similarly, because there is no explicit time shift, the reader assumes that the event is temporally continuous with the previous one. Furthermore, the reader is likely to make the causal inference that John and Bill were prompted to look around by the scream, and that the scream was produced by the child that was nearly drowning. The third sentence foregrounds John, who is already in the discourse model. Here, the reader is likely to infer from John's action that he has the intention to save the child – a goal that is accomplished in the final sentence. Together, these interactions between the text and the reader's knowledge generate a representation in long-term memory in which the events are firmly interconnected.

Whereas situation model theory is specific about the macro-level of event representation, it does not address the micro-level, that is, the internal course of the events in question. At the time these theories were developed, the building blocks of the situation model were assumed to be propositional representations with, for instance, protagonists in a story represented as 'tokens' (Zwaan & Radvansky, 1998) or pointers to information stored in long-term memory. However, research since then has shown that sentence comprehension involves performing a mental simulation of the described events (e.g., Glenberg & Kaschak, 2002; Zwaan, Stanfield, & Yaxley, 2002). A hallmark of these mental simulations is that they offer a solution to the *grounding problem* (Harnad, 1990): for linguistic symbols to have any meaning at all, they must at some point map onto their referents in the world. Simulations accomplish this by drawing on the comprehender's perceptual and motor memory traces of past interactions with the world, which can be combined to covertly reenact the described events. Take the sentence *John did not hesitate*

and dove into the water, which can elicit activation in multiple modalities. For a first-person action perspective, one might recruit one's own motor program for diving. Alternatively, one could take an observer perspective, perhaps that of Bill, and simulate what it would be like to see a person dive into a canal.

The evidence for simulation of perceptual properties of described events is robust (Zwaan & Pecher, 2012) and it is hard to deny the involvement of the motor system in language processing (Glenberg & Gallese, 2011), although its functional role is subject to debate (Fischer & Zwaan, 2008; Mahon & Caramazza, 2008). However, the majority of studies have been concerned with language comprehension at the sentence level, which meant that participants were presented with large sets of sentences that had no meaningful connection to one another. Obviously, this is not an accurate resemblance of how individuals typically use language. Unfortunately, very little research has been done with regard to mental simulation embedded in discourse. This might have left us with a skewed picture of its role in naturalistic language processing (see Clark, 1997, for a detailed treatment of how ignoring facts about language use may misguide theory). The present research aims to assess the role of grounded event representations in discourse comprehension.

Contrasting predictions about the occurrence of grounded event representations in connected discourse can be made. On the one hand, a narrative context enables comprehenders to construct detailed mental representations. Our image of John would arguably have been richer if we had previously read that he was tall, dark-haired, and athletic (Albrecht & O'Brien, 1993). Conversely, we would not have had a strong cue to simulate where the child was drowning if we had not been told about the canal. According to the *spatial grounding hypothesis* (Beveridge & Pickering, 2013), a situation model in which the spatial relations between entities are specified facilitates the simulation of described events. So, an 'immersed' experience (Mar & Oatley, 2008; Zwaan, 2004) is more likely to take place in a story than in isolated sentences. Consistent with this idea, listening to short passages elicited stronger activation in brain areas associated with processing visual motion than listening to individual words (Dravida, Saxe, & Bedny, 2013), and coherent passages about concrete actions elicited stronger activation in auditory-specific and motor-specific areas than scrambled passages made up of the same sentences (Kurby & Zacks, 2013).

On the other hand, discourse presents constraints on what is represented in detail, and to what extent modality-specific systems are recruited. Whereas hearing words like *kick* and *throw* in isolation may activate the 'leg' and 'hand' areas, respectively (Hauk, Johnsrude, & Pulvermüller, 2004), this picture changes when we consider language comprehension in context. For instance, Dutch action verbs preceded by a literal sentence context (e.g., 'Iedereen was blij toen oma de taart *aansneed*', literally 'Everyone was happy when grandma the cake *sliced'*) elicited greater motor activity than action verbs in

a non-literal context (e.g., 'Iedereen was blij toen oma een ander onderwerp *aansneed*', literally 'Everyone was happy when grandma a different topic *sliced'*, with 'sliced' meaning 'broached') (Schuil, Smits, & Zwaan, 2013). Similarly, simply stated actions within a story implying left- or rightward manual rotation (e.g., *He started the car*) elicited motor resonance, but planned actions (e.g., *He wanted to start the car*) did not (Zwaan, Taylor, & de Boer, 2010).

Extensive discourse may impose even more stringent limitations on event-internal simulations than minimal sentence contexts. If the beginning of our story had dealt with John's preparations for a diving contest, the act of diving would be highly relevant. The actual context, however, renders John's manner of entering the water tangential to the plot, perhaps allowing the reader to take a 'shortcut' by leaving a fine-grained representation of this particular event out of the situation model. One reason comprehenders would do so, we suggest, lies in the attentional resources that building event representations and maintaining coherence necessarily require. Although both constitute highly practiced skills in proficient adults, they cost *some* processing capacity and time (Madden & Zwaan, 2006). The latter may be a particular bottleneck during online auditory comprehension. Thus, a strong focus on creating grounded event representations may interfere with coherence-building, and vice versa. All the same, a coherent situation model to a certain extent depends on grounded event representations. For instance, it has been found that a visual secondary task (i.e., remembering a dot pattern) interferes more with coherence monitoring than a verbal secondary task (i.e., remembering a digit string) (Fincher-Kiefer, 2002). We may take this to mean that perceptual systems are necessarily involved in comprehending individual events, and that lacking a sufficiently grounded representation, establishing coherence becomes difficult. The question, then, is just how much of grounded representation is needed (see also Zwaan, in press).

Our approach to this question is as follows: if a potential behavior that benefits a coherent global representation conflicts with a potential behavior that benefits an event-internal representation, which do comprehenders choose? If grounding the ongoing event in perceptual representations is necessary, then comprehenders are expected to prioritize those processes that instantiate local meaning over more coherence-based processes. A way to test this is by tracking listeners' preferential inspection of two complementary aspects of linguistically described scenes using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In our study, we presented participants with spoken sentences like *The doctor was tired and knelt in order to tie his laces* concurrently with visual scenes in which one picture matched the agent, but not the action (a standing doctor), and one picture matched the action, but not the agent (a kneeling baker). All sentences were embedded in short narratives, such that each character had been introduced before. If at a given point in time one picture has a greater 'pull' on

visual attention than others, it can be inferred that it is most pertinent to the listener. So, if listeners prioritize a grounded representation of the described events, they will predominantly look at the picture representing the action; if listeners prioritize constructing a coherent situation model, they will predominantly look at the picture representing the agent.

A potential problem in this setup is the fact that in a typical English sentence the noun denoting an agent comes before the verb denoting an action. Therefore, if we find a preference for inspecting the picture corresponding to the agent, this might be a consequence of it being mentioned first, in combination with a participant's averseness to produce additional eye movements. So, to reveal a genuine preference for one of the pictures, word order should be ruled out as a confounding factor. The Dutch language provides a convenient solution, because both subject-verb and verb-subject word orders can be used to create felicitous sentences.

Hypotheses

Two findings in the eye-tracking literature are important to mention here. First, when there is a likely referential candidate, nouns and verbs alike can guide eye movements across specific elements of a scene, and both depictions of actions and persons serve as likely referential candidates (e.g., Knoeferle & Crocker, 2006; Kukona, Fang, Aicher, Chen, & Magnuson, 2011). Second, eye movements are driven by dynamically updated mental representations, rather than the visual or linguistic stimulus per se (Altmann & Kamide, 2009). So, we can reasonably expect listeners to fixate elements of a scene that are not the current focus of the linguistic input, and a sufficiently strong mental representation to 'override' the immediately available visual context.

Our hypotheses concern the listener's gaze behavior from the moment the speaker utters the word (either a noun or a verb) that initiates the mismatch with the visual display. At this point, assuming that the listener is looking at the picture referred to in the first part of the sentence and does not know whether the sentence will mismatch the visual scene (because there was a 50:50 distribution of matching and mismatching trials), we can anticipate at least four more or less plausible scenarios, which are summarized in Figure 1.

Under the *no updating* hypothesis, participants will keep looking at the picture that was referred to in the first part of the sentence, regardless of whether it represents the character or the action. We allow for this possibility in light of evidence that language users do not consistently revise early sentence interpretations (Ferreira, Bailey, & Ferraro, 2002) and fail to respond to semantic anomalies (Barton & Sanford, 1993). By contrast, under the *immediacy dominance* hypothesis, which is derived from basic accounts of incremental processing (e.g., Carpenter & Just, 1983), participants will look at the picture that best matches the incoming information (see also Kukona & Tabor, 2011). This entails a switch from the character to the action during noun-first sentences, and a switch from

the action to the character during verb-first sentences. Under the *action dominance* hypothesis, participants will predominantly inspect the action described in the sentence, even when the incoming information points to a different picture. This scenario would be consistent with a strong event-internal focus. Under the *agent dominance* hypothesis, we expect the opposite pattern, with participants looking at the picture corresponding to the agent of the sentence, even when new information disqualifies this picture as the single best available representation. This would argue against a strong event-internal focus and emphasize the role of coherence at a global level.

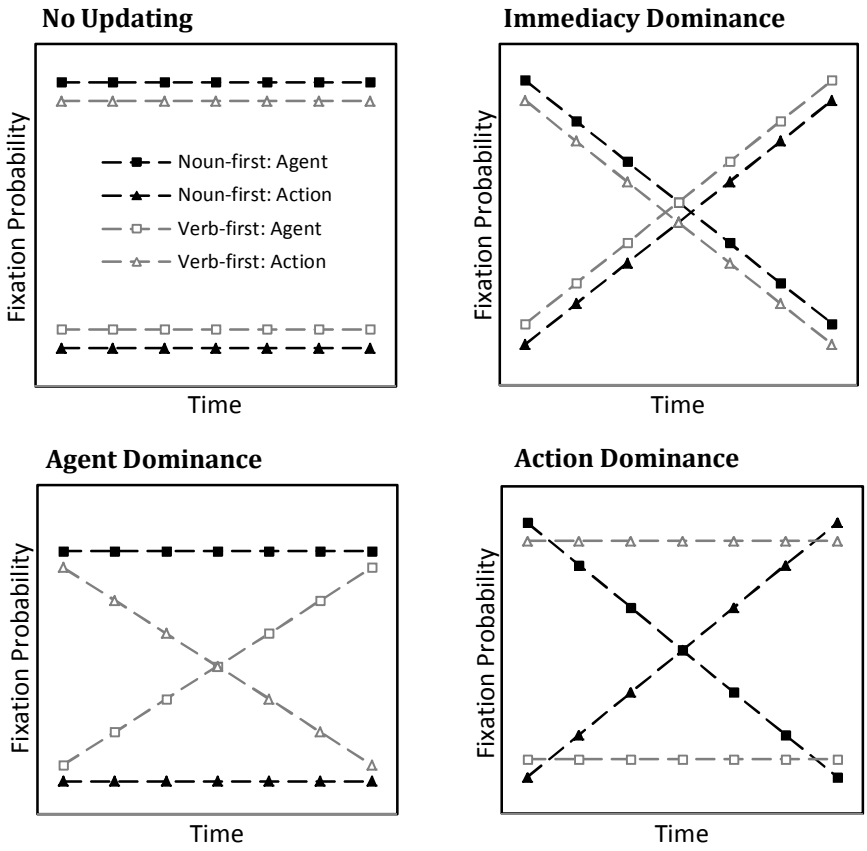


Figure 1. Predicted fixation probabilities for the character and action pictures under the four hypotheses.

To distinguish these scenarios, we used growth curve analysis (GCA), which permits the treatment of time as a continuous variable (Mirman, Dixon, & Magnuson, 2008). Because none of the pictures in a given experimental trial matches both the action and the agent described in the sentence, it is unlikely for the comprehension system to

settle on one as an ideal representation. Rather, a certain amount of switching between pictures may be expected as the trial proceeds. GCA is a suitable method for capturing these changes in gaze state over time. Other techniques, such as analysis of variance on fixation proportions that are aggregated over time, do not provide such information, unless we would select smaller time windows. Because we did neither have strong assumptions about the shape of the fixation probability distributions – the linear trends in Figure 1 are a simplification – nor about the timing and duration of participants' preference for one of the pictures, we could only select these time windows arbitrarily or post-hoc. GCA enables us to model the shape of the multinomial probability distribution throughout the mismatch window with orthogonal polynomials. The focus of our analysis is not on the specific parameters of this model, but rather on the overall pattern of fixations on the agent and action picture and its resemblance to the patterns in Figure 1. We therefore took a two-stage approach, in which we first estimated a series of polynomial power functions to describe the complex fixation patterns as fully and accurately as possible, and compared the graphs resulting from the most parsimonious model with the hypothesized scenarios (Experiment 1). To investigate the validity of this interpretation, we used the confidence intervals around the estimations, and a systematic comparison with fixations on other AOIs and fixations during trials that did not contain a mismatch. Then, to evaluate the stability of these results, we conducted a confirmatory replication of the experiment with a new sample of participants (Experiment 2).

Experiment 1

Method

Participants. Forty Dutch-speaking university students (31 females, ages 18-28, mean age = 20.33) participated in the experiment as a requirement of the educational program. All had normal or corrected vision.

Materials.

Linguistic stimuli. Sixteen experimental stories were constructed, consisting of five sentences each. Sentence 1 introduced a location and two protagonists. Sentences 2 and 3 elaborated the situation, without mentioning either of the protagonists specifically. In sentence 4, one of the protagonists performed an action. Sentence 5 wrapped up the story. In the verb-first word order, sentence 4 started with a temporal or locational adverb (e.g., 'immediately', 'there'), which is a way of licensing verb-fronting in Dutch. To be able to observe eye movements between the noun and the verb in sentence 4, they were temporally separated by an attributive adjective in the verb-first word order (e.g., *kneeled the tired doctor*) and a predicative adjective in the noun-first word order (e.g., *the doctor was tired and kneeled*). The adjectives (e.g., 'brave', 'happy') were selected to not have strong implications for the way the action was performed. Table 1 gives a sample story with both word orders. In addition, 16 filler stories were constructed. These had the same

structure as the experimental stories, but were presented concurrently with a visual display that matched the event that was described in the target sentence (i.e., one of the pictures was consistent with both the subject and the verb). The purpose of these stories was to make sure that participants could not predict whether a given trial would contain a mismatch.

The stories were recorded by a female native speaker of Dutch. The time between onset of the noun and the verb was kept as constant as possible, although we had to allow for some variability. The mean difference was 1465 ms (range 940-2420 ms) for noun-first sentences and 1134 ms (ranges 880-1950 ms) for verb-first sentences. For noun-first and verb-first versions of the same sentence, the difference was on average 332 ms for noun-first sentences (range –230-790 ms). In all pairs but one the difference was larger for the noun-first sentence.

Table 1
Sample Story (Translated from Dutch)

The doctor and the monk were taking a long hike.
It was a hot day and the trail was pretty difficult.
After an hour they took a break in a shaded area.
Noun-first: The doctor was tired and kneeled in order to tie his laces.
Verb-first: There kneeled the tired doctor in order to tie his laces.
Then he took a sandwich from his backpack and sat down.

Visual stimuli. Forty grey-scale line drawings were created for the purpose of the experiment. There were eight characters that were all identifiable by their profession (a hunter, a doctor, a gardener, a police officer, a monk, a soldier, a baker, and an athlete). There were five drawings of each character: one in a neutral pose, i.e., standing upright, and four performing distinct bodily actions (e.g., jumping, crouching, and pointing). There were 16 unique actions, with each action being depicted by two characters.

The relationship between the stories and the visual display (see Figure 2 for an example) was as follows: in experimental trials, the display showed the character that was the subject of sentence 4 in a neutral pose (‘agent’), a character that was not mentioned in the story performing the action described in sentence 4 (‘action’), the other character that was mentioned in sentence 1 in a neutral pose (‘unfocused’), and an extraneous character performing an unrelated action (‘extraneous’). In filler trials, the display showed the ‘unfocused’ and ‘extraneous’ characters, along with the character that was mentioned in sentence 4 performing the action described in that sentence (‘target’), and an extraneous character in a neutral pose (‘extraneous-neutral’). Each unique action drawing was used twice across the experiment: once as ‘action’ and once as ‘extraneous’. Furthermore, each unique non-action drawing was used eight times: twice as ‘agent’ (in experimental trials

only), twice as 'extraneous-neutral' (in filler trials only) and four times as 'unfocused'. The assignment of characters to stories and visual displays was fixed.

Each drawing occupied an area of approximately 400 x 400 pixels against a white background of 1600 x 1200 pixels. The pictures were placed in a square in the center of the screen, such that the horizontal and vertical distance between the centers of any two adjacent pictures was 600 pixels. The assignment of drawings to quadrants was randomized.

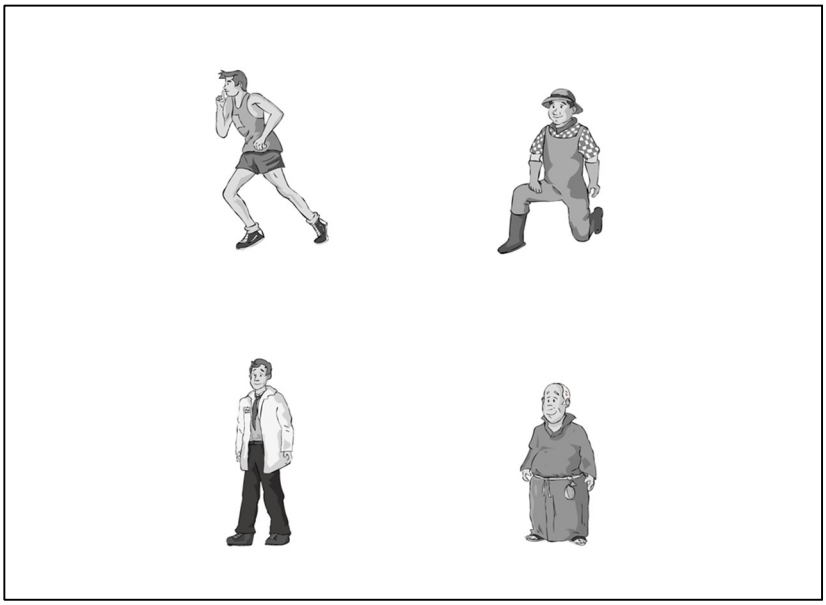


Figure 2. Visual world display corresponding to the sample story in Table 1. The athlete is the 'extraneous' picture, the gardener is 'action' picture, the doctor is the 'agent' picture, and the monk is the 'unfocused' picture.

Comprehension questions. All 32 stories were followed by a four-alternative comprehension question. One set of 16 questions ('picture questions') asked which protagonist performed the action in the target sentence (e.g, *Who knelt to tie his laces?*). The question was presented as text in the upper half of the screen, with a horizontal 'line-up' of the four characters in a neutral pose below. Participants were instructed to click on the picture corresponding to the answer. Another 16 questions ('text questions') asked for a fact that was mentioned at some point in the story (e.g., *Why did the monk and the soldier arrive late?*). The answer options were four text boxes that could be clicked. Ordering of answer options was randomized. We were mainly interested in the answers to the pictorial questions for experimental items, but included the textual questions to ensure that participants paid attention to the stories in their entirety and could not predict the question. To avoid a one-to-one correspondence between story type and question type,

four experimental stories were followed by a textual question and four filler stories by a pictorial question.

Design. The experiment had a one-way factorial design with word order of the target sentence (noun-first vs. verb-first) as a within-subjects variable. We created two lists to counterbalance the assignment of stories to word order condition. Trials were presented in random order without constraints on the alteration between experimental and filler stories.

Procedure. Participants were seated approximately 60 cm from the eye-tracker, a Tobii 2150 system with a sampling rate of 50 Hz. The experiment started with a five-point calibration of the eye-tracker. Participants then read instructions that they were going to listen to a set of short narratives while viewing pictures, and that each narrative would be followed by a question. They were asked to keep looking at the display throughout the task and to answer the questions by using the information they had heard.

Before the first trial, the eight characters were introduced successively for three seconds with their professions (by which they were referenced throughout the experiment) printed below. Each trial started with a fixation cross that was displayed in the center of the screen for four seconds before the onset of the first sentence. The story was presented over stereo speakers with intervals of 800 ms between sentences. The fixation cross remained until the offset of sentence 3, upon which it was replaced by the visual world display. This yielded 800 ms of preview time before the onset of the critical sentence. The visual world display remained until the offset of sentence 5, upon which it was replaced by the question and the four answers. As soon as the participant had selected an answer, the next trial started. The entire experiment lasted approximately 25 minutes.

Data analysis. Eye movement data from 200 ms (corresponding to the first time point where fixations could be driven by the mismatch) to 3700 ms (corresponding to the offset of the shortest sentence) after the onset of the mismatch were modeled using GCA with orthogonal power polynomials. The model consisted of two hierarchically related levels, with level-1 parameters describing the effect of time and level-2 describing these parameters in terms of population means, fixed effects, and random effects (Mirman et al., 2008). The equation for level-1 is defined as follows:

$$\pi^{s}_{ijt} = \beta_0{}^{s}_{ij} + \beta_1{}^{s}_{ij}\, Time + \beta_2{}^{s}_{ij}\, Time^2 + \ldots, \beta_p{}^{s}_{ij}\, Time^p + \mu^{s}_{ijt}.$$

The dependent variable on level-1, π, is a multinomial variable with five categories[1], indicated by the superscript *s, s=1, …, S*, representing the five discrete states (i.e., fixating

[1] The use of a multinomial variable is distinct from a regular GCA approach in visual world studies, in which separate models are fit for each AOI or AOI is included as a categorical predictor in a single overall model. The former approach precludes a direct comparison between AOIs, and the latter assumes independency of observations, while in fact fixations on different AOIs within a display are negatively correlated, potentially

one of the four pictures or the background) a participant *i, i=1, …, n,* could be in at a given time *t*, *t=1*, …, *T*, on a specific trial *j*, *j=1*, …, *J*. The predictor $\beta_{0}^{s}{}_{ij}$ represents the intercept of participant *i* at trial *j*, and predictors $\beta_{1}^{s}{}_{ij}$ to $\beta_{p}^{s}{}_{ij}$ are the polynomial terms, *p=1*, …, *P*, representing the linear, quadratic, and *p*th-order effects of time, and are allowed to vary across individuals and trials. The error term, μ^{s}_{ijt}, is assumed to be normally distributed around zero, and is allowed to vary across participants and trials. The predictors $\beta_{0}^{s}{}_{ij}$ to $\beta_{p}^{s}{}_{ij}$ form the dependent variables on level-2, for which the equations are defined as follows:

$$\beta_{0}^{s}{}_{ij} = \gamma_{00}^{s} + \gamma_{00}^{s}{}_{i} + \gamma_{01}^{s} WO + \gamma_{01}^{s}{}_{i} WO + \varepsilon_{0}^{s}{}_{ij}.$$
$$\beta_{p}^{s}{}_{ij} = \eta_{p0}^{s} + \eta_{p0}^{s}{}_{i} + \eta_{p1}^{s} WO + \eta_{p1}^{s}{}_{i} WO + \xi_{p}^{s}{}_{ij}.$$

The level-2 submodel for $\beta_{0}^{s}{}_{ij}$ consists of a fixed intercept γ_{00}^{s} and fixed effect for word order γ_{01}^{s}, and a random intercept $\gamma_{00}^{s}{}_{i}$ and random effect for word order $\gamma_{01}^{s}{}_{i}$, which are allowed to vary for each participant *i*. Error term $\varepsilon_{0}^{s}{}_{ij}$ is assumed to be normally distributed around zero. Similarly, each of the level-2 submodels for $\beta_{p}^{s}{}_{ij}$ represent the population average polynomial effects of the level-1 model, consisting of a fixed and random intercept η_{p0}^{s} and $\eta_{p0}^{s}{}_{i}$ and a fixed and random effect for word order η_{p1}, and $\eta_{p1}^{s}{}_{i}$. Error term $\xi_{p}^{s}{}_{ij}$ is assumed to be normally distributed around zero. The error terms ε and ξ are allowed to correlate.

To establish what number of time polynomials best described the data, we fitted a series of models. First, we estimated a baseline model with only a fixed intercept and a random intercept for participants, called Model 1. In Model 2 we included an effect for word order on the intercept. In Models 3-10, we added submodels for the linear, quadratic, cubic, quartic, quintic, sextic, septic, and octic polynomials, respectively, each including an effect of word order. We used the difference between these nested models in –2 * log likelihood ($-2LL_{diff}$) to assess the effect of expanding the model. This deviance statistic has a χ^2 distribution, with the degrees of freedom equal to the difference in the number of parameters between each pair of nested models. Analyses were performed in the program Latent GOLD 5.0 (Vermunt & Magidson, 2013).

Results

Comprehension accuracy. Participants showed few memory errors: mean accuracy was 97.3% (*SD* = 5.1%) for picture questions corresponding to experimental stories and 98.8% (*SD* = 5.5%) for picture questions corresponding to filler stories. With

leading to spurious results. The present model takes this interdependency into account, affording not only a comparison of fixation probabilities for the same AOI across conditions, but also for different AOIs within and across conditions.

regard to textual questions, mean accuracy was 89.4% (*SD* = 13.7%) for experimental stories and 90.6% (*SD* = 8.1%) for filler stories. Because the small amount of variance in accuracy on the theoretically interesting picture questions did not warrant a comparison between the viewing preferences of individuals with different patterns of memory errors, we did not analyze this relationship further.

Eye movements. We grouped fixation data into bins of 100 ms (comprising five samples each). Up to two consecutive bins that were coded as missing due to tracking loss were set to the value of their surrounding bins when these had identical AOI values. Given that it takes approximately 200 ms to plan and execute a saccadic eye movement (Rayner, Slowiaczek, Clifton, & Bertera, 1983), it is reasonable to assume that the eye had not moved to another AOI and back during that time. Trials with more than 900 ms of tracking loss (2.50 % of all trials, including fillers) were removed from the analysis. After this recoding procedure, 1.94% of all data bins constituted missing values.

Figure 3a plots the observed fixation probabilities for the agent and action pictures in different word orders in experimental trials and for the target picture in filler trials. At the beginning of a given trial, participants were most likely to inspect the action and target picture. This is plausible, because both were more visually salient than the agent picture (see Figure 2). When the speaker uttered the first word that could be related to the visual display, participants shifted their gaze to the corresponding picture. The initial rate at which they did so was equal for action, agent, and target pictures, but fixations on target and action pictures started to decrease 1000 ms after the onset of the verb in verb-first sentences. (These trends were not part of the original hypotheses, but are analyzed in more detail in Appendix B). This means that after processing the verb, but not the noun, participants started inspecting other parts of the display. Indeed, as the intercept values in Figure 5 show, participants were more likely to inspect each of the other AOIs in verb-first sentences than in noun-first sentences. This leads to a baseline difference at the onset of the mismatch window: while in both word orders there was a clear advantage for the just-referenced picture, this advantage was greater in noun-first trials. So, the assumed initial fixation probabilities in Figure 1 are not entirely warranted.

Furthermore, at the onset of the mismatch, for trials with the same word order, there were no clear differences between filler trials and experimental trials, suggesting that participants initially inspected both types of display in a similar fashion, as expected. Finally, although we did not model the data from filler trials during the mismatch window, the likelihood of fixating the partly matching agent picture in experimental trials appears only slightly smaller than that of fixating the completely matching target picture in filler trials, and equal after 1500 ms.

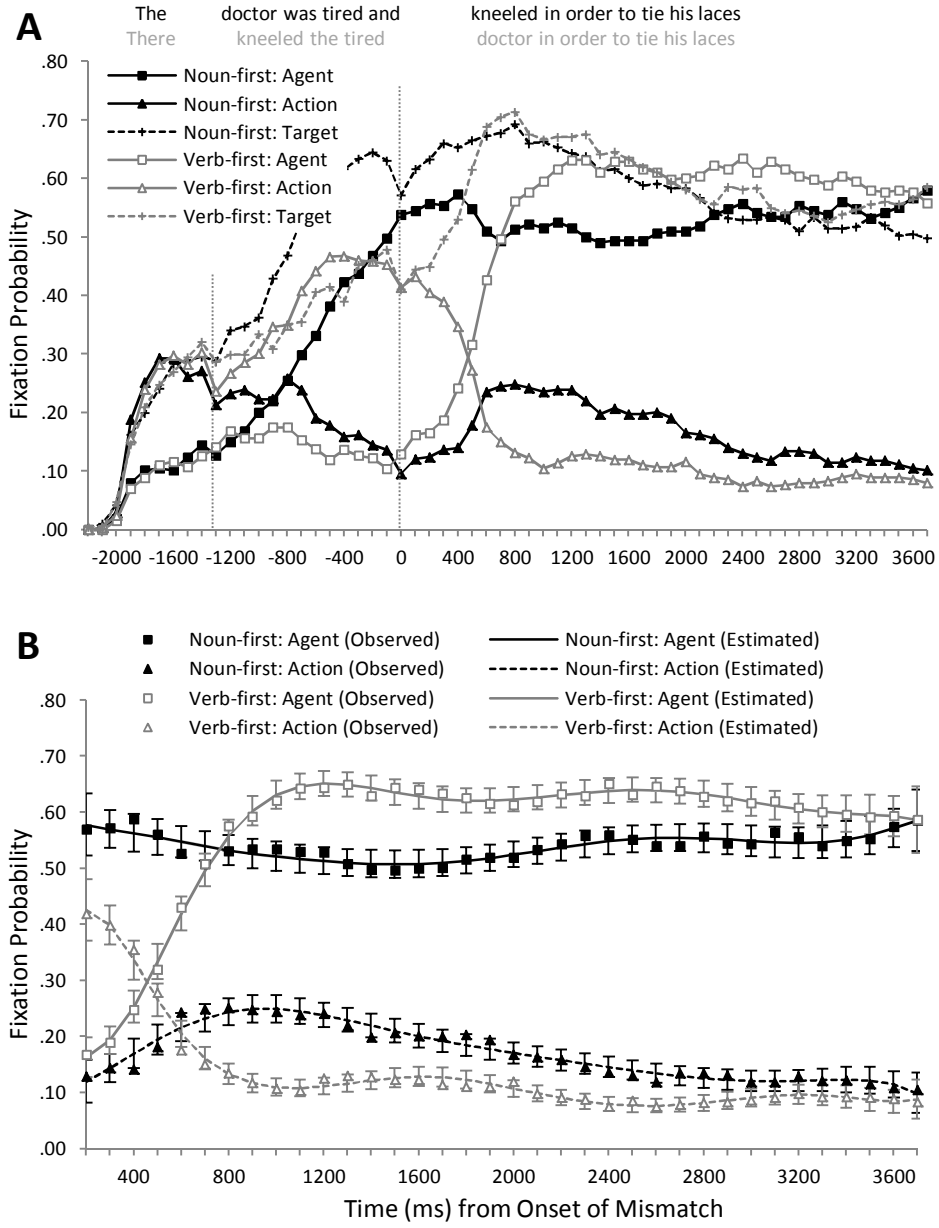


Figure 3. Panel A: Fixation probabilities for character, action, and filler target pictures in Experiment 1. The time between the onset of the noun and the verb (indicated by the vertical lines) was variable across trials, so the probabilities between time points –500 and 0 were based on fewer observations; as a compromise, we plotted all probabilities that were based on at least 100 observations per condition and from there 'stitched' the graph together with the graph at time point 0. Panel B: Observed and estimated fixation probabilities for the character and action picture during the mismatch window. Error bars represent 95% CIs around the estimated probabilities.

Growth curve analysis. Table 2 gives the fit statistics for the inclusion of each set of predictors. The model with a seventh-order polynomial function described the data most parsimoniously and we therefore chose to interpret this model.[2] As can be observed in Figure 3b, this model captures the observed fixations well, following each major bend in the curves. The parameter estimates and statistical tests for the polynomial terms for all AOIs are provided in Appendix A. To determine which scenario, if any, the data support, we present a detailed description of the estimated fixation probabilities for the agent and the action picture, using the 95% CIs as plotted in Figure 3b for statistical inference where appropriate.

Table 2
Summary of the Growth Curve Analyses for Experiments 1 and 2

Experiment	Model	Description	n_{par}	*LL*	$-2LL_{diff}$	*p*
1	1	Intercept	8	-27265.700	-	-
	2	Previous model + WO	16	-27217.713	95.975	<.001
	3	Previous model + Time * WO	33	-26680.840	1073.745	<.001
	4	Previous model + $Time^2$ * WO	49	-26485.204	391.273	<.001
	5	Previous model + $Time^3$ * WO	65	-26362.516	245.375	<.001
	6	Previous model + $Time^4$ * WO	81	-26298.277	128.478	<.001
	7	Previous model + $Time^5$ * WO	97	-26268.901	58.753	<.001
	8	Previous model + $Time^6$ * WO	113	-26248.437	40.927	.001
	9	Previous model + $Time^7$ * WO	129	-26232.181	32.513	.009
	10	Previous model + $Time^8$ * WO	145	-26224.759	14.843	.536
2	1	Intercept	8	-24798.961	-	-
	2	Previous model + WO	16	-24756.644	84.635	<.001
	3	Previous model + Time * WO	33	-24399.808	713.671	<.001
	4	Previous model + $Time^2$ * WO	49	-24275.642	248.333	<.001
	5	Previous model + $Time^3$ * WO	65	-24205.630	140.024	<.001
	6	Previous model + $Time^4$ * WO	81	-24175.555	60.150	<.001
	7	Previous model + $Time^5$ * WO	97	-24158.084	34.943	.004
	8	Previous model + $Time^6$ * WO	113	-24141.053	34.062	.005
	9	Previous model + $Time^7$ * WO	129	-24110.072	61.962	<.001
	10	Previous model + $Time^8$ * WO	145	-24101.727	16.688	.406

[2] This is a more complex model than common in visual world studies, where polynomial functions are often restricted to quadratic or cubic. The reason we needed higher-order polynomials is that we modeled eye movements during a protracted interval (3500 ms), so naturally there are more bends in the curve. Moreover, we modeled fixations on five different AOIs, so there is a larger chance that at least one distribution has a complex shape.

The strongest transitions in gaze state probability occurred in verb-first trials. The probability of fixating the action picture steeply decreased, reaching its lowest point at approximately 1100 ms and then remaining low (below .13, CI [.11 .14], which is similar to the values of the other AOIs, as displayed in Figure 5). The probability of fixating the agent picture steeply increased, reaching its highest point at approximately 1300 ms and then remaining high (above .58, CI [.53 .65]).

Changes in gaze state probabilities were less pronounced in noun-first trials. The probability of fixating the action picture slightly increased, reaching its highest point at approximately 1000 ms and then returning to its initial level (below .13, CI [.10 .14]). The probability of fixating the agent picture slightly decreased, reaching its lowest point at approximately 1400 ms and then returning to its initial level (above .54, CI [.52 .57]).

Across word orders, from 500 ms onwards the agent picture was reliably more likely to be fixated than the action picture. This difference increased until approximately 1100 ms and then remained constant throughout the rest of the analysis window. While this pattern provides clear support for the agent dominance scenario, it appears to be modulated by bottom-up information processing, such that the most recently referenced picture enjoyed a small and temporary advantage over the one referenced in the first part of the sentence: participants were reliably more likely to fixate the action picture in noun-first trials than in verb-first trials between 700 and 2900 ms, and reliably more likely to fixate the agent picture in noun-first trials between 900 and 3200 ms.

Discussion

We anticipated four possible patterns of gaze behavior in response the mismatch between the visual display and the linguistic stimulus: *no updating*, *immediacy dominance*, *agent dominance*, and *action dominance*. The results show a strong prevalence of looks toward the subject of the target sentence, thus supporting the agent dominance scenario. In line with this, participants virtually performed at ceiling level on the memory questions, showing no intrusion of the action-related information. However, both results may be caused by strategic viewing behavior: to answer the memory questions correctly, it would in half of all trials suffice to (verbally) remember the subject of the target sentence or (visually) its picture, despite the purposeful variation in questions. Thus, looks toward the agent picture might reflect preparation for the memory question rather than a natural preference. To rule out this task-induced bias, we used more varied comprehension questions in Experiment 2.

Furthermore, the results show a short-lived increase of fixations on the most recently referenced picture, suggesting that incremental processing of the linguistic stimulus also affected eye movements, though not as profoundly as the agent dominance. This interaction between agent dominance and immediacy dominance, although apparently reliable, was not predicted a priori. So, besides to validate our general

approach, Experiment 2 also serves to confirm this particular finding. If it emerges again in a new sample, then we may be confident that it reflects a genuine effect.

Experiment 2

Method

Participants. Thirty-five Dutch-speaking university students (24 females, ages 17-27, mean age = 20.17) participated in the experiment as a requirement of the educational program. None had participated in Experiment 1. All had normal or corrected vision.

Materials, Design, and Procedure. Experiment 2 was identical to Experiment 1, except for the comprehension questions. All 32 stories were followed by a textual question, with six questions asking who performed an action in the target sentence (e.g., *Who was the first to jump over the puddle?*), six questions asking what action was performed in the target sentence (e.g., *What did the hunter do when he saw the lady being robbed?*), and the other questions asking about a detail mentioned in one of the other sentences (e.g., *Why did the monk and the soldier arrive late?*). Accuracy was still expected to be close to ceiling and therefore dropped as a dependent variable of theoretical interest, but these questions should ensure that participants listened for comprehension.

Results

Comprehension accuracy. Mean accuracy was 85.4% (*SD* = 18.9%) for experimental stories and 96.8% (*SD* = 6.2%) for filler stories. The lower accuracy compared to Experiment 1 was presumably due to the questions being less predictable and more often referring to less accessible information from the first sentences of the story.

Eye movements. We made the a priori decision to use the same data inclusion and recoding plan as in Experiment 1. After recoding, 3.18% (including filler trials) constituted missing values. The observed fixation probabilities are plotted in Figure 4a. The major trends prior to the mismatch window resembled those of Experiment 1. However, the values of the strongest attractors, namely the agent picture in experimental trials and the target picture in filler trials, were considerably lower, while the values of the action picture were similar. This means that other AOIs attracted more attention than in Experiment 1. The values of these AOIs at the intercept of Figure 5 suggest that participants were most likely to fixate the background. As a consequence, the baseline advantage for the agent picture over the action picture at the onset of the mismatch window was smaller than in Experiment 1.

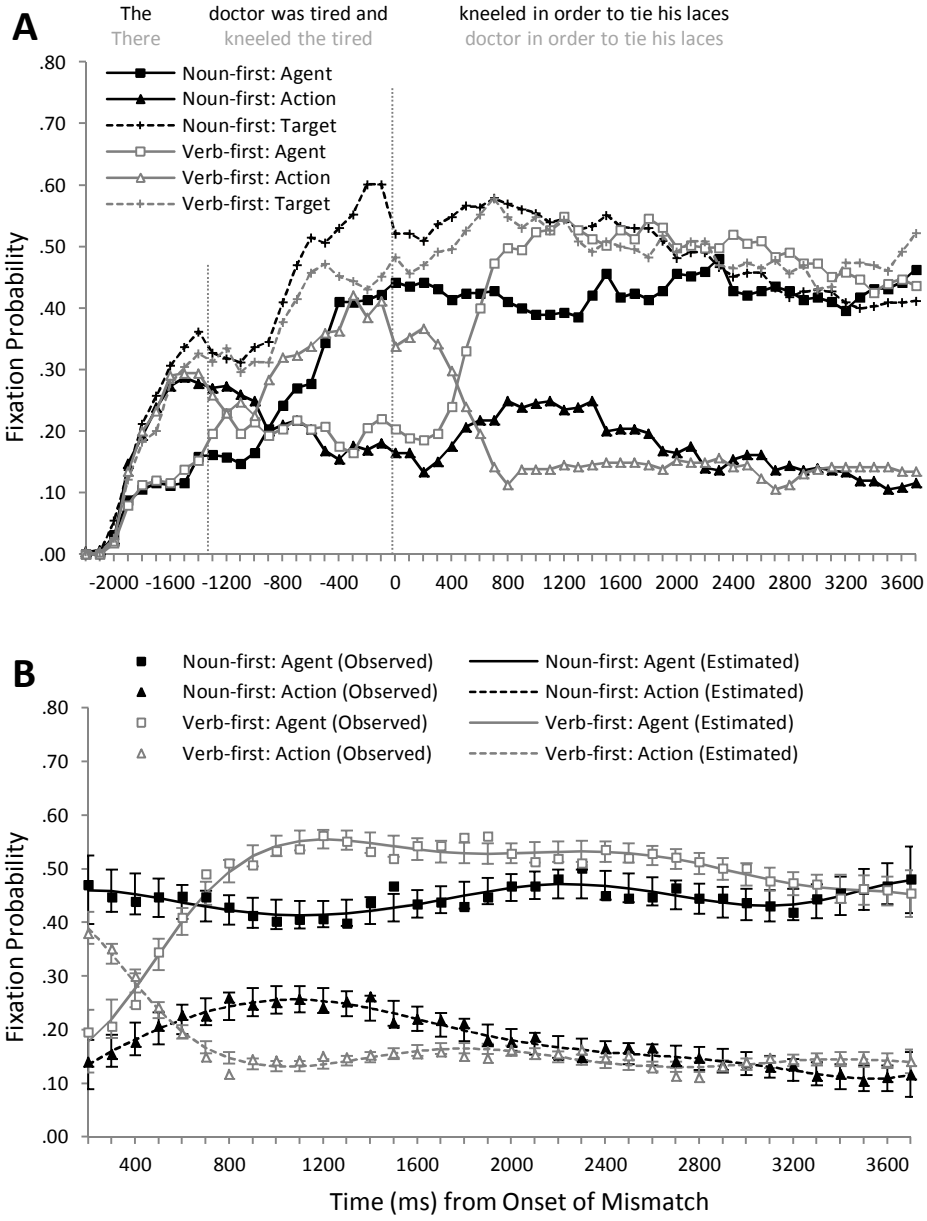


Figure 4. Panel A: Fixation probabilities for character, action, and filler target pictures in Experiment 2. The time between the onset of the noun and the verb (indicated by the vertical lines) was variable across trials, so the probabilities between time points –500 and 0 were based on fewer observations; as a compromise, we plotted all probabilities that were based on at least 100 observations per condition and from there 'stitched' the graph together with the graph at time point 0. Panel B: Observed and estimated fixation probabilities for the character and action picture during the mismatch window. Error bars represent 95% CIs around the estimated probabilities.

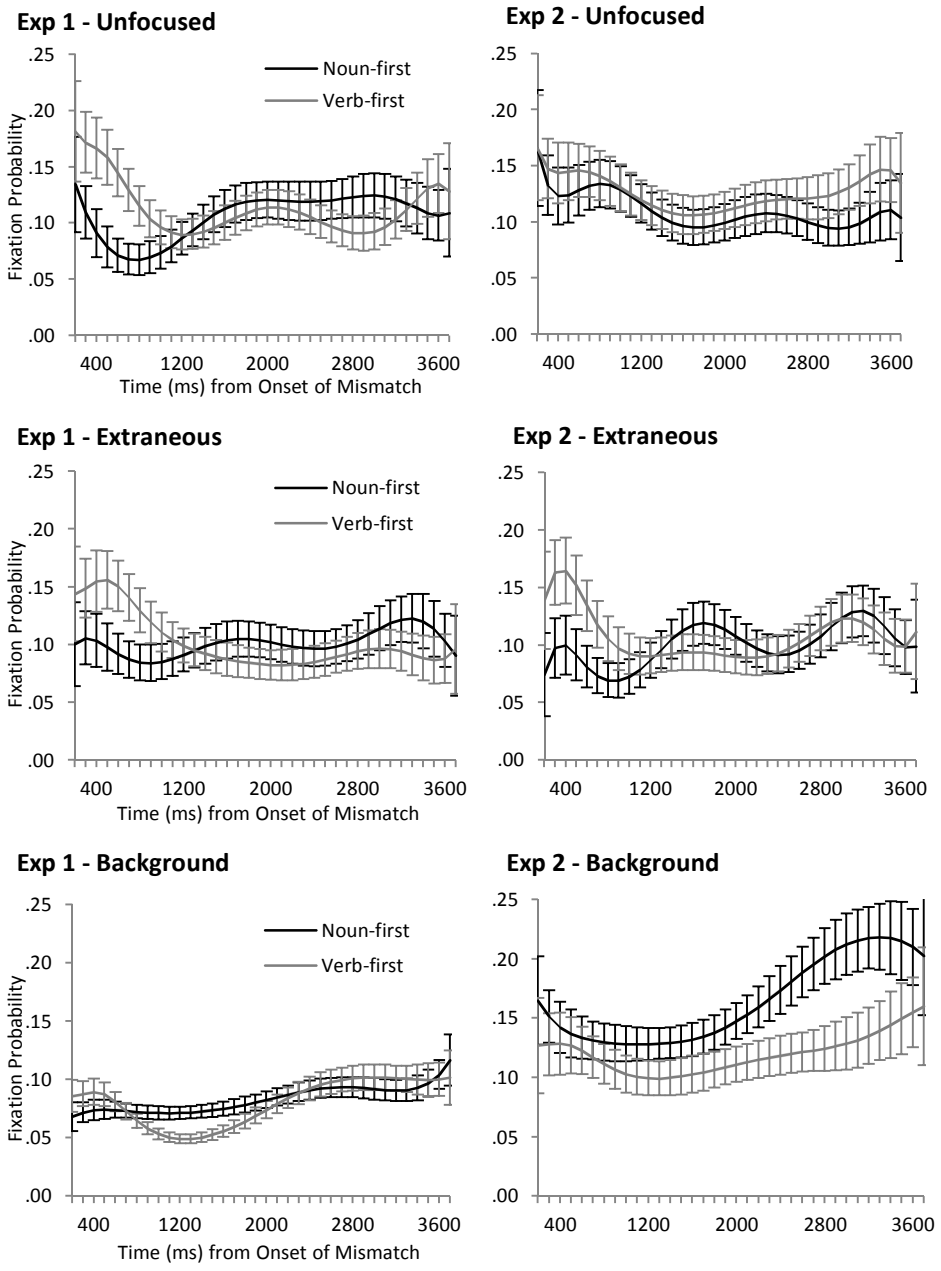


Figure 5. Estimated fixation probabilities for the unfocused and extraneous picture and the background in Experiments 1 and 2. Error bars represent 95% CIs around the estimated probabilities.

Growth curve analysis. Table 2 gives the fit statistics for the inclusion of each set of predictors. As in Experiment 1, the model with a seventh-order polynomial function provided the best fit while meeting the parsimoniousness criterion. We therefore interpret the estimated fixation probabilities resulting from this model, focusing on the agent and action pictures. The parameter estimates and statistical tests for all AOIs are provided in Appendix A.

The strongest transitions in gaze state probability occurred in verb-first trials. The probability of fixating the action picture steeply decreased, reaching its lowest point at approximately 1000 ms and then remaining low (below .17, CI [.15 .18]). The probability of fixating the agent picture steeply increased, reaching its highest point at approximately 1300 ms and then remaining high (above .45, CI [.39 .51]).

Changes in gaze state probabilities were less pronounced in noun-first trials. The probability of fixating the action picture slightly increased, reaching its highest point at approximately 900 ms and then returning to its initial level (below .12, CI [.07 .16]). The probability of fixating the agent picture slightly decreased, reaching its lowest point at approximately 1100 ms and then returning to its initial level (above .43, CI [.40 .46]).

Across word orders, from 500 ms onwards participants fixated the agent picture reliably more than the action picture. This preference increased until approximately 1200 ms and then remained constant throughout the rest of the analysis window, supporting the agent dominance scenario. Also, we found an additive effect of incremental processing, such that the newly referenced picture enjoyed an increase in fixation probability. This effect was of shorter duration than in Experiment 1, with the action picture in noun-first trials showing a reliable advantage over verb-first trials between 700 and 1800 ms, and the character picture in verb-first trials showing an advantage over noun-first trials between 800 and 3000 ms.

Discussion

The results of Experiment 2 resembled those of Experiment 1 in at least three respects: first, before the sentence started to mismatch the visual display, participants fixated the aspect of the scene that the speaker referred to. This tendency was stronger for noun-first trials than for verb-first trials. Second, during the mismatch window, there was a marked preference for the agent picture across word orders. Third, this preference was modulated by bottom-up processing, as indicated by a minor increase in fixations on the newly referenced aspect. The results also deviated from Experiment 1 in that the probability of fixating the agent picture throughout the analysis window was lower. It is plausible that this difference was caused by the fact that focusing on the agent picture was no longer helpful in answering the questions. Furthermore, this difference meant that participants spent more time looking at the background, rather than the other pictures. Looks toward the background might constitute a disengagement from the environment

(Glenberg, Schroeder, & Robertson, 1998), so as to reduce the interference from visual stimuli on the processing verbal information. Apart from these minor differences, Experiment 2 underlines the stability and the validity of the findings of Experiment 1.

Importantly, the agent dominance could not be explained by the demands of the memory task. However, an asymmetry in the design was that each unique 'neutral' picture was used eight times across trials, whereas any given 'active' picture was only used twice. It could be that previously seen pictures were more likely to be fixated because of their familiarity. However, familiarity alone cannot explain these results, because the 'unfocused' picture occurred equally often as the agent picture, yet was approximately five times less likely to be fixated, and was equally likely to be fixated as the 'extraneous' picture, even though it occurred four times as often. Moreover, any type of frequency bias could only come into being during the experiment. Cursive inspection of the aggregated eye movement data from the first two experimental trials suggested that before such a bias could reasonably exist, the patterns corresponding to the agent dominance scenario were present.

General Discussion

Two visual world experiments investigated the relative importance of global coherence and detailed event-internal representations during language comprehension. In both experiments, only one depicted aspect of a described event could be overtly attended to at any given moment. Tracking listeners' gaze over time as a sentence unfolded allowed us to infer *which* particular aspect would be most strongly activated *when*. Both experiments led to the same finding: comprehenders focused on the agent involved in an event rather than the action. Importantly, this preference was only marginally affected by word order. This provides clear support for the agent dominance scenario. We now go on to outline two potential explanations for this observation.

First, we suggest that the macro-level of representation ranks higher in importance than the micro-level when the reader or listener is engaged in comprehending connected discourse. As a result, behavior that supports a coherent situation model tends to be prioritized over behavior that supports comprehension of individual events (similar to why readers make bridging inferences online, but not elaborative inferences; see Graesser, Singer, & Trabasso, 1994). How does looking at the character picture help building a coherent mental model? One reason is that other situational dimensions, particularly intentionality, are tightly connected to the protagonists that populate the discourse. For example, when John from the opening paragraph dives into the water, the reader is likely to infer that he does so in order to rescue the child. This information subsequently serves as an organizing structure for the story, until the goal is completed (Lutz & Radvansky, 1997). As such, it is crucial that this particular goal is attributed to John, and not to another character, lest the comprehender end up with an inaccurate or

incoherent mental model. To the extent that such information (be it inferred or explicitly stated) would automatically be mapped onto the representation of the character that is in the center of visual attention, looks toward an inappropriate character should be avoided. Conversely, looks toward the appropriate character might serve to better connect the inferred goal to the character in memory.

By the same token, looks toward the character picture may have served to facilitate comprehension of the upcoming discourse. All stories continued after the target sentence, often (in 11 out of 16 cases) by referring back to its subject with a pronoun (e.g., *Then he took a sandwich from his backpack and sat down*). In case participants came to expect such continuation, looks toward the agent picture might reflect anticipation of another event involving the same agent. One may argue that the demand for comprehension beyond the current sentence is an artifact of the length of our stories. But 'looking forward' is precisely what readers and listeners do when processing connected discourse (Gordon, Grosz, & Gilliom, 1993). Corpus studies (e.g., Givón, 1983) have shown that sentence subjects that are realized as a definite noun phrase tend to be continued for at least one more sentence – a fact known as *topic persistence*. We see no reason to assume that the prediction mechanisms involved in language comprehension, for which there is ample evidence within and across words (see Altmann & Mirković, 2009, for a review), should stop at sentence boundaries. So, rather than the present study creating artificial task demands, previous studies using isolated sentences might inadvertently have created an artificial *lack* of task demands.

The above explanation is broadly consistent with the idea that eye movements across external space can act as a temporary extension of working memory (Ballard, Hayhoe, Pook, & Rao, 1997). So, looking at the agent picture might aid comprehension not because it contributes useful visual input, but because it serves as a deictic pointer to store and retrieve information (Altmann, 2004; Johansson & Johansson, 2014; Richardson & Spivey, 2000). However, if one assumes that listeners *do* use the visual cues provided by the environment to scaffold comprehension, there is an explanation for the agent dominance that does not require the notion of coherence at the level of discourse: it might be easier to look at the standing doctor and use this visual information to simulate this person kneeling than to look at the kneeling gardener and simulate a different person performing that action. The former would likely involve a reconfiguration of already present visual information (e.g., a character's arms and legs), akin to mental rotation (Parsons, 1987; Shepard & Metzler, 1971), while the latter would require a more complicated substitution process, in which typical attributes (e.g., a character's clothes and face) are suppressed and replaced. So, if listeners engage in mentally transforming a given percept in the service of comprehension, then looking at the character may be the most efficient choice.

Because participants showed high accuracy on all questions, including those about the actions described in the target sentence, it appears that the representation of action did not suffer, despite the fact that in 31.4% of all trials the action picture was never fixated during the mismatch window (as opposed to 3.8% for the agent picture). Admittedly, due to the short delay, the memory task may have been too easy to reveal subtle differences in representation strength, and might have tapped a verbatim memory trace of the target sentence (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Nevertheless, this state of affairs lets us speculate about an account in the spirit of grounded cognition, in which the perceptual and the motor system cooperated in processing a given target sentence. Specifically, while the perceptual system extracted task-relevant information from the environment (i.e., the identity of the agent), the motor system contributed by covertly executing the action. More research is necessary to test this multi-modal comprehension mechanism, which necessarily involves the coordination of multiple (i.e., internal and external) perspectives.

Note that we have framed the discussion around where participants chose to look, rather than where they chose *not* to look. Another way of interpreting the data would be that they are about what type of mismatch listeners were most tolerant to. However, such an interpretation is problematic in light of the high fixation probability of the agent picture, which in both experiments was equal to that of the target picture in filler trials. Moreover, participants were explicitly told they were free to look where they wanted on the screen. If their eye movements would have mainly been driven by mismatch avoidance, we would have expected lower fixation rates.

Finally, participants' gaze behavior prior to the mismatch prompts some further considerations. Specifically, verbs in Experiment 1 were associated with more transient visual attention than nouns (although the analyses in Appendix B suggest that the statistical evidence for this trend was weak): while participants initially shifted their gaze to the first picture that the speaker referenced, they were more likely to shift it away again in verb-first sentences. At the same time, fixations on the agent picture in noun-first sentences continued to rise. This may not explain the full magnitude of the subsequent agent dominance, but it does suggest that nouns and verbs differ in their baseline impact on attention to visual scenes. An important question, then, is whether this difference should be attributed to the intrinsic properties of nouns versus verbs (see Gentner, 2006), or to the pictures that were linked with them. The present data cannot provide a clear answer to that question. However, it is tempting to draw a parallel between the transient nature of that what many verbs denote – an action – as opposed to nouns, which denote an entity that remains constant across time, and participants' fixations on the corresponding pictures. More generally, this seems to align with the deeply entrenched bias in children and adults to map language onto constant entities (Gillette, Gleitman, Gleitman, &

Lederer, 1999). Of course, it is their constancy that is also fundamental to the importance of characters in stories, providing a basis for coherence.

The present study is one of the first to investigate the interplay between grounded event representation and the construction of coherent situation models. Moreover, it adds to our understanding of situated language processing by showing that listeners selectively attend to a visual representation depending on whether it matches the agent described in a concurrent utterance, regardless of whether it matches the described action. While the experimental separation between actions and agents may not have a direct analog in everyday communicative situations, the findings have implications for situations in which language only partly overlaps with the visual context. Examples are a discussion between two persons in a room about where everybody was standing at last week's party, or a child reading a book about a character involved in a series of adventurous events with just a single picture on each page.

A few issues remain open for further research. For instance, all events were described in the simple past. This grammatical form often receives a perfective interpretation, which highlights the resultant state of a given event (Ferretti, Kutas, & McRae, 2007). Perhaps a grammatical form that cues the listener to construe the ongoing event in more detail (e.g., progressive aspect) would have elicited more attention to the action picture. Furthermore, although our linguistic stimuli were conceived as coherent units of connected discourse, they are still a long way from extensive narrative discourse, being so-called 'textoids' (Graesser, Millis, & Zwaan, 1997). To what extent our results generalize to longer texts, with more developed characters and more intricate plots, remains open to investigation. We predict that the coherence-driven agent bias will be even more pronounced when individual events play a smaller role in the overall story.

To conclude, the present work provides a systematic demonstration of the experience during listening or reading that rather than imagining in detail how events described in the narrative take shape, it is important to represent at least the characters that were involved in them and that might be affected. So, simulating how John dives into the water may not be all that essential. What we want to know is what John does next and whether his attempt at rescue will be successful. Perhaps this makes that real-time comprehension at times involves less extensive grounded representation than theories of embodied cognition have previously suggested (see also Zwaan, in press).

Appendix A
Parameter Estimates and Statistical Tests for Growth Curve Analyses

Table A1
Parameter Estimates and Statistical Tests for the Fixed Effects

		Experiment 1				Experiment 2			
AOI *s*	Predictor	Est	*SE*	Wald *Z*	*p*	Est	*SE*	Wald *Z*	*p*
Agent	Intercept γ_{00}^{1}	1.380	.014	98.290	<.001	.956	.016	61.478	<.001
Action	Intercept γ_{00}^{2}	.176	.019	9.372	<.001	.146	.018	8.300	<.001
Unfocused	Intercept γ_{00}^{3}	-.064	.020	-3.185	.002	-.259	.021	-12.635	<.001
Extraneous	Intercept γ_{00}^{4}	-.146	.021	-7.043	<.001	-.346	.021	-16.554	<.001
Background	Intercept γ_{00}^{5}	-1.346	.034	-39.867	<.001	-.497	.026	-19.414	<.001
Agent	WO γ_{01}^{1}	-.030	.014	-2.211	.027	-.003	.014	-.202	.840
Action	WO γ_{01}^{2}	.161	.019	8.659	<.001	.060	.018	3.421	.001
Unfocused	WO γ_{01}^{3}	-.051	.020	-2.577	.010	-.091	.021	-4.411	<.001
Extraneous	WO γ_{01}^{4}	-.007	.020	-.346	.730	-.003	.021	-.136	.890
Background	WO γ_{01}^{5}	-.073	.033	-2.208	.027	.037	.024	1.521	.130
Agent	Time η_{10}^{1}	.403	.095	4.243	<.001	.288	.094	3.075	.002
Action	Time η_{10}^{2}	-1.779	.107	-16.567	<.001	-1.197	.104	-11.553	<.001
Unfocused	Time η_{10}^{3}	-.125	.115	-1.080	.280	-.709	.132	-5.359	<.001
Extraneous	Time η_{10}^{4}	-.524	.119	-4.421	<.001	.090	.123	.730	.470
Background	Time η_{10}^{5}	2.024	.164	12.353	<.001	1.528	.147	10.401	<.001
Agent	Time * WO η_{11}^{1}	-.714	.078	-9.123	<.001	-.389	.082	-4.770	<.001
Action	Time * WO η_{11}^{2}	.258	.106	2.431	.015	-.302	.104	-2.912	.004
Unfocused	Time * WO η_{11}^{3}	.360	.114	3.150	.002	-.233	.118	-1.972	.049
Extraneous	Time * WO η_{11}^{4}	.468	.116	4.031	<.001	.403	.122	3.292	.001
Background	Time * WO η_{11}^{5}	-.372	.156	-2.389	.017	.521	.139	3.758	<.001
Agent	Time^2 η_{20}^{1}	-.753	.076	-9.982	<.001	-.773	.083	-9.315	<.001
Action	Time^2 η_{20}^{2}	.135	.107	1.265	.210	-.155	.104	-1.494	.140
Unfocused	Time^2 η_{20}^{3}	.083	.111	.750	.450	.158	.118	1.332	.180
Extraneous	Time^2 η_{20}^{4}	.206	.119	1.731	.083	.163	.124	1.321	.190
Background	Time^2 η_{20}^{5}	.330	.156	2.115	.034	.607	.143	4.233	<.001
Agent	Time^2 * WO η_{21}^{1}	1.073	.076	14.163	<.001	.810	.082	9.823	<.001
Action	Time^2 * WO η_{21}^{2}	-.661	.108	-6.132	<.001	-.591	.104	-5.710	<.001
Unfocused	Time^2 * WO η_{21}^{3}	-.290	.112	-2.605	.009	.017	.119	.140	.890
Extraneous	Time^2 * WO η_{21}^{4}	.046	.118	.390	.700	-.230	.123	-1.871	.061
Background	Time^2 * WO η_{21}^{5}	-.168	.156	-1.075	.280	-.005	.142	-.036	.970
Agent	Time^3 η_{30}^{1}	.678	.076	8.866	<.001	.403	.082	4.917	<.001
Action	Time^3 η_{30}^{2}	.349	.111	3.146	.002	.047	.105	.452	.650
Unfocused	Time^3 η_{30}^{3}	-.171	.114	-1.498	.130	.295	.120	2.457	.014
Extraneous	Time^3 η_{30}^{4}	.068	.119	.568	.570	-.297	.123	-2.408	.016
Background	Time^3 η_{30}^{5}	-.923	.169	-5.479	<.001	-.449	.142	-3.150	.002
Agent	Time^3 * WO η_{31}^{1}	-.614	.077	-7.983	<.001	-.433	.083	-5.220	<.001
Action	Time^3 * WO η_{31}^{2}	.574	.112	5.147	<.001	.561	.105	5.345	<.001
Unfocused	Time^3 * WO η_{31}^{3}	-.340	.114	-2.984	.003	.065	.120	.541	.590
Extraneous	Time^3 * WO η_{31}^{4}	-.112	.120	-.935	.350	.098	.124	.792	.430

		Experiment 1				Experiment 2			
AOI *s*	Predictor	Est	*SE*	Wald *Z*	*p*	Est	*SE*	Wald *Z*	*p*
Background	Time[3] * WO η_{31}^{5}	.492	.166	2.958	.003	-.291	.145	-2.012	.044
Agent	Time4 η_{40}^{1}	-.326	.075	-4.332	<.001	-.088	.082	-1.079	.280
Action	Time[4] η_{40}^{2}	-.066	.107	-.619	.540	.129	.104	1.245	.210
Unfocused	Time[4] η_{40}^{3}	.469	.114	4.116	<.001	.199	.122	1.635	.100
Extraneous	Time[4] η_{40}^{4}	-.216	.118	-1.829	.067	-.040	.124	-.321	.750
Background	Time[4] η_{40}^{5}	.139	.160	.872	.380	-.200	.143	-1.398	.160
Agent	Time[4] * WO η_{41}^{1}	.276	.076	3.620	<.001	.338	.082	-4.003	<.001
Action	Time[4] * WO η_{41}^{2}	-.459	.108	-4.246	<.001	-.416	.104	1.058	<.001
Unfocused	Time[4] * WO η_{41}^{3}	.062	.113	.544	.590	.128	.121	.693	.290
Extraneous	Time[4] * WO η_{41}^{4}	.060	.118	.511	.610	.086	.124	-.959	.490
Background	Time[4] * WO η_{41}^{5}	.061	.160	.380	.700	-.136	.142	-4.147	.340
Agent	Time[5] η_{50}^{1}	.020	.075	.267	.790	.124	.082	1.513	.130
Action	Time[5] η_{50}^{2}	-.336	.107	-3.156	.002	-.106	.104	-1.023	.310
Unfocused	Time[5] η_{50}^{3}	-.163	.114	-1.428	.150	.055	.120	.454	.650
Extraneous	Time[5] η_{50}^{4}	-.189	.118	-1.604	.110	-.126	.123	-1.028	.300
Background	Time[5] η_{50}^{5}	.668	.160	4.168	<.001	.054	.141	.382	.700
Agent	Time[5] * WO η_{51}^{1}	.113	.075	1.506	.130	.001	.081	.011	.990
Action	Time[5] * WO η_{51}^{2}	.358	.107	3.339	.001	.129	.104	1.243	.210
Unfocused	Time[5] * WO η_{51}^{3}	-.048	.114	-.419	.680	-.033	.119	-.276	.780
Extraneous	Time[5] * WO η_{51}^{4}	-.191	.120	-1.596	.110	-.275	.124	-2.221	.026
Background	Time[5] * WO η_{51}^{5}	-.232	.160	-1.451	.150	.178	.144	1.240	.210
Agent	Time[6] η_{60}^{1}	.259	.074	3.484	.001	.057	.081	.706	.480
Action	Time[6] η_{60}^{2}	-.086	.106	-.811	.420	-.026	.103	-.256	.800
Unfocused	Time[6] η_{60}^{3}	.004	.113	.036	.970	.094	.121	.775	.440
Extraneous	Time[6] η_{60}^{4}	.096	.118	.814	.420	-.100	.124	-.805	.420
Background	Time[6] η_{60}^{5}	-.273	.159	-1.722	.085	-.025	.141	-.176	.860
Agent	Time[6] * WO η_{61}^{1}	-.145	.074	-1.946	.052	-.134	.081	-1.655	.098
Action	Time[6] * WO η_{61}^{2}	.082	.106	.770	.440	.186	.103	1.807	.071
Unfocused	Time[6] * WO η_{61}^{3}	.173	.114	1.514	.130	.007	.120	.057	.950
Extraneous	Time[6] * WO η_{61}^{4}	-.268	.118	-2.281	.023	-.143	.122	-1.170	.240
Background	Time[6] * WO η_{61}^{5}	.158	.157	1.006	.310	.084	.141	.597	.550
Agent	Time[7] η_{70}^{1}	-.167	.073	-2.271	.023	-.115	.080	-1.436	.150
Action	Time[7] η_{70}^{2}	.083	.103	.811	.420	.109	.102	1.078	.280
Unfocused	Time[7] η_{70}^{3}	-.042	.111	-.373	.710	-.507	.117	-4.339	<.001
Extraneous	Time[7] η_{70}^{4}	.094	.115	.815	.420	.441	.122	3.623	<.001
Background	Time[7] η_{70}^{5}	.031	.151	.206	.840	.072	.138	.518	.600
Agent	Time[7] * WO η_{71}^{1}	.062	.073	.842	.400	.067	.080	.829	.410
Action	Time[7] * WO η_{71}^{2}	-.285	.103	-2.772	.006	.007	.102	.068	.950
Unfocused	Time[7] * WO η_{71}^{3}	.136	.112	1.215	.220	-.122	.117	-1.044	.300
Extraneous	Time[7] * WO η_{71}^{4}	.010	.115	.085	.930	.188	.122	1.534	.120
Background	Time[7] * WO η_{71}^{5}	.078	.151	.516	.610	-.139	.139	-1.004	.320

Note. Effect coding is used for Word Order (Noun-first: 1, Verb-first: -1).

Appendix B
Analysis of Gaze Data Prior to the Mismatch Between the Linguistic Stimulus and the Visual Display

In Experiment 1 we noticed unanticipated differences in how nouns and verbs guided visual attention across the display. In particular, it seems that fixations on the action picture triggered by the verb started to decline earlier than fixations on the agent picture triggered by the noun. To assess whether these differences were reliable, we fitted a growth curve model using the procedure described in the section *Data analysis* under Experiment 1. The input data were all fixations between 200 ms (the moment at which the first language-driven eye movements could be expected) and 1200 ms after the onset of the first word in the sentence. Because the time between the onset of the noun and the verb was variable across trials, this window represents a compromise. On the one hand, using the length of the shortest trial as a cutoff would mean discarding all data beyond 800 ms, making this analysis obsolete. On the other hand, using the length of the longest trial would result in estimations for the time points near the end of the window being based on few observations, potentially leading to unreliable estimates. Therefore, we used all time points for which we had at least 100 observations in each word order.

The fit statistics for the inclusion of each set of parameters are given in Table B1. The model with a quadratic polynomial was the most parsimonious solution for Experiment 1. For Experiment 2, this was the model with a linear polynomial. The estimated graphs for Experiments 1 and 2 are displayed in Figure B1. In Experiment 1, the probability distribution for the action picture in verb-first sentences shows a clear quadratic trend, with an inflection point at 1000 ms, but at no point were there reliably fewer fixations compared to the agent picture in noun-first sentences. In Experiment 2, there was no indication that fixations on the action picture started to decline before the mismatch. Rather, the distribution for the action picture had a shallower slope than that of the agent picture. Again, there was no point at which there were reliably fewer fixations on the action than on the agent.

From this mixed evidence we conclude that there is no strong statistical support for the notion that fixations on the action picture started to decline prior to the mismatch. More likely, the baseline differences at the onset of the mismatch window are caused by a stronger overall increase of fixations on the agent picture. Adding to this, the average delay between the onset of the first word and the mismatch was longer in noun-first sentences. So, participants might have had more time to fixate the appropriate picture in these trials.

Table B1

Summary of the Growth Curve Analyses for Prior to the Onset of the Mismatch in Experiments 1 and 2.

Experiment	Model	Description	n_{par}	*LL*	$-2LL_{diff}$	*p*
1	1	Intercept	8	-8997.301	-	-
	2	Previous model + WO	16	-8765.533	463.535	<.001
	3	Previous model + Time * WO	33	-8585.029	361.007	<.001
	4	Previous model + Time^2 * WO	49	-8564.192	41.675	<.001
	5	Previous model + Time^3 * WO	65	-8554.626	19.133	.262
2	1	Intercept	8	-7800.144	-	-
	2	Previous model + WO	16	-7722.830	154.627	<.001
	3	Previous model + Time * WO	33	-7573.614	298.434	<.001
	4	Previous model + Time^2 * WO	49	-7562.020	23.188	.109
	5	Previous model + Time^3 * WO	65	-7559.481	5.078	.995

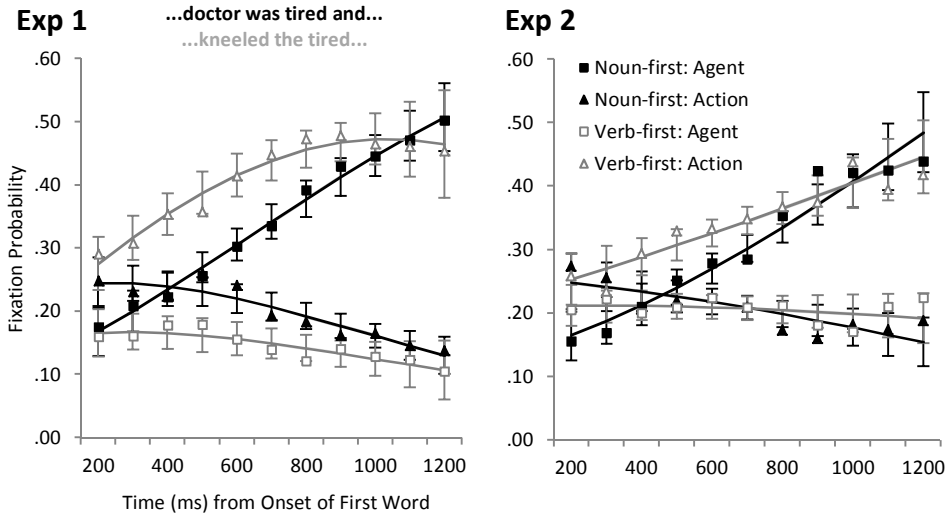


Figure B1. Estimated and observed fixation probabilities following the onset of the first word in the sentence. Error bars represent 95% CIs around the estimated probabilities.

Chapter 5

Does Picture Orientation Constrain Spatial Situation Models?*

* This chapter has been submitted for publication as Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (submitted). Does picture orientation constrain spatial situation models?

Abstract

Does seeing a pictorial illustration prior to reading influence how readers mentally represent aspects of a situation other than those that were depicted? In four experiments, we investigate whether the orientation of a contextual picture of a character and the direction described in a sentence influence where readers focus their attention in the visual field. Contrary to our predictions, we found no evidence that response times in a picture verification task with objects presented left or right on a computer screen were affected by the interplay between visual and linguistic cues. In a meta-analysis, the effect of this interaction was close to zero. However, responses were reliably faster after sentences describing movement of the character toward an object, indicating that comprehenders may have adopted a character-internal perspective. These results suggest that if readers construct perceptual simulations of spatial relations, the perspective from which they do so is not affected by recent visual experiences.

Does the presence of a pictorial illustration influence how the reader construes the events described in a given text? In a classic experiment, Bransford and Johnson (1972) had participants listen to sentences like *If the balloons popped, the sound wouldn't be able to carry since everything would be too far away from the correct floor*, which they were later asked to recall. While participants generally performed poorly, they did much better when they had previously seen a picture of a man playing a serenade below a woman's window high up in an apartment building, with the guitar connected to an amplifier held aloft by balloons. The authors suggested that the memory trace of the visual context helped participants to integrate the sentences into a coherent mental representation.

However, the above example represents a rather unusual case, because the text is difficult to understand without the picture, as evidenced by comprehensibility ratings from the same experiment, and even with the picture does not convey a very plausible situation. While a large literature documents the benefits of pictorial illustrations when their role in providing meaning to a passage of text is less critical (see Carney & Levin, 2002, for a review), an important question is precisely how they interact with the comprehension process. According to theories of embodied cognition, which posit that language comprehension is grounded in sensorimotor experiences (Barsalou, 1999; Zwaan, 2004), it is quite likely that illustrations have a direct impact. Recently seen pictures constitute a highly salient and accessible class of sensorimotor experiences and might influence the reader's mental representation in a number of ways.

For instance, participants who had viewed a set of pictures that contained, among other things, a toothbrush, read the prepositional phrase *in the cup* in the sentence *Aunt Karen finally found the toothbrush in the cup beside the mirror* faster when this toothbrush had been vertically oriented than when it had been horizontally oriented (Wassenburg & Zwaan, 2010). This suggests that comprehenders immediately recruit a memory trace of a recent visual experience when they process a related linguistic description. In this case, there was a straightforward mapping between the picture and the noun. But would a visual memory trace also constrain mental simulations of other entities in a described situation? This question is the focus of the present research. In particular, we investigate whether the orientation of a contextual picture of the agent of a sentence (e.g., a doctor facing either left or right) influences the way readers simulate other objects that are described in relation to this agent (e.g., *The doctor walked toward the cabinet where he kept the patient's file*). If the memory trace of a particular illustration serves as an 'anchor' for the situation model, this would give us new insights into how visual memory interacts with linguistic processing.

Compelling evidence that mental representations retain analog spatial properties comes from research on eye movements. When looking at a blank screen, listeners unwittingly shift their gaze according to the directions described in short passages, such as upward when hearing *On the 10th floor, a woman is hanging her laundry out the window.*

On the 29th floor, two kids are sitting on the fire escape smoking cigarettes (Spivey & Geng, 2001). Eye movements also reflect more fine-grained spatial relations, such as *at the center, to the far right,* and *in front of* (Johansson, Holsanova, & Holmqvist, 2006). In another experiment by these authors, participants studied a picture and later described this picture from memory. Gaze locations concurrent with naming of specific elements closely resembled the locations of these elements in the original stimulus. Thus, eye movements seem to coordinate elements of a mental model with elements of the visual field.

Similarly, listening to sentences that convey movement along a horizontal (e.g., *the miner pushes the cart*) or vertical axis (e.g., *the ship sinks in the ocean*) selectively interferes with visual discrimination of stimuli on the corresponding axis (Richardson, Spivey, McRae, & Barsalou, 2003). This suggests that language comprehension recruits specific parts of the visual system which then cannot, or less efficiently, be used for a secondary task. Building on this, Bergen, Lindsay, Matlock and Narayanan (2007) found that listening to short statements conveying up- or downward movement (e.g., *the mule climbed*) hampered visual performance in the corresponding half of the visual field. So, language-induced imagery does not only encode the axis along which movement takes place, but also the specific part of the axis with which a particular motion event is associated.

Can a contextual picture modulate these imagery effects? In the present study, participants read spatial sentences such as *The doctor walked toward the cabinet where he kept the patient's file*. Assuming that walking will invoke a horizontal image schema, the destination (i.e., the cabinet) may be located either left or right on the horizontal axis. Evidence suggests that by default comprehenders represent the trajectory as going from left to right – a phenomenon known as *spatial agency bias* (Chatterjee, Southwood, & Basilico, 1999; Maass & Russo, 2003) – and thus imagine the destination on the right. However, it has been suggested that the spatial agency bias is susceptible to contextual manipulation (Suitner & Giacomantonio, 2012). Now suppose that we read this sentence in the context of a picture of the agent (i.e., the doctor), which happened to be facing left. Would this change the imagined location of the desk on the horizontal axis?

This question goes beyond previous research in that we ask if language users attend to locations in the visual field where they might mentally simulate an object, not based on linguistic information alone, but on the integration of linguistic and contextual visual information. A related topic has been explored by Altmann and Kamide (2009). In one of their experiments participants viewed clipart scenes, including one that contained pictures of a woman, a wine glass and a bottle, and an empty table. Concurrently, they heard either *The woman will put the glass on the table* or *The woman is too lazy to put the glass on the table,* both of which were continued by *she will carefully pour the wine into the glass.* During this second segment, with the scene unchanged, participants tended to fixate the empty table in the *will put*-condition, and the glass in the *too lazy to put-*

condition. This suggests that participants' eye movements were driven by their updated mental representation as encoded from language, rather than by the immediate visual context. This provides valuable information about the dynamic interplay between language and vision. However, because the target location was indexed to a visible object, the listeners' visual attention was strongly determined by the particular context that the researchers provided. As such, this work does not allow us to conclude whether comprehenders would also simulate a target location if it had not been present – which is an important question if one wants to know if this process also occurs during language comprehension in non-deictic situations.

To summarize, the evidence reviewed suggests that pictures have a strong potential of constraining mental models, which in turn reflect some of the properties of the perceptual world. Against this background, it seems reasonable to hypothesize that comprehenders use contextual pictures to ground the spatial relations in described scenes.

The Present Study

In all experiments described in this paper, participants were briefly presented with a line drawing of a person or vehicle facing left or right. Subsequently, they read a sentence in which this person or vehicle was described as moving toward or away from an object. Then they saw a line drawing of an object for which they had to indicate whether or not it had been mentioned in the sentence. Crucially, this object was displayed either on the left or the right half of the screen – thus matching or mismatching the spatial location implied by the combination of the contextual picture and the sentence.

Based on the literature review, we predict that comprehenders will simulate the target or the source location with reference to the contextual picture, and be faster to verify the object if it is presented to the front of the character in *toward*-sentences and to the back of the character in *away*-sentences. This hypothesis would be supported by a three-way interaction between character orientation (left vs. right), direction (away vs. toward), and object location (left vs. right). Although some studies mentioned in the introduction found language comprehension to interfere with visual categorization, we expected facilitation for two reasons. First, the objects shown were those described in the sentence, resulting in a higher degree of integratability (Kosslyn, 1994). Second, we used a longer inter-stimulus interval, so that participants were more likely to have completed a mental simulation of the events by the time they had to give a response (see also Kaschak et al., 2004).

Alternatively, it might be possible that participants perform the task without being influenced by the contextual picture, but still imagine the described objects in specific locations of the visual field. In this case, responses should follow a spatial agency bias and be faster for an object on the left if the sentence describes motion away from it, and faster for an object on the right if the sentence describes motion toward it. Finally, in

case the spatial relations described in the sentences do not lead participants to differentially allocate their attention across the visual field – as would be predicted by a purely propositional view of language comprehension – reaction times should be equal across conditions.

All experiments reported in this paper were presented online in the Qualtrics research suite (http://www.qualtrics.com). Participants were adult US residents recruited through Amazon's Mechanical Turk (http://www.mturk.com). After each experiment we asked participants 1) to guess what the purpose of the study was, 2) to indicate the amount of distraction and level of noise in their environment (on a 9-point scale), 3) what type of monitor they used, and 4) some demographical questions (age, gender, level of education, native language).

We used the same pre-determined inclusion criteria for participants and trials in each experiment. Response times for each experiment were analyzed in a hierarchical linear regression model with orientation, direction, and location and all their higher-order interactions as fixed effects. The model also included a by-participant random intercept and by-participant random slopes for orientation, direction, and location and all their higher-order interactions (see Barr, Levy, Scheepers, & Tily, 2013, for a justification of this maximal random effect structure). The statistical analyses were conducted with the software LatentGold 5.0 (Vermunt & Magidson, 2013).

Experiment 1

Method

Participants. We recruited 120 participants, of which 118 (41 males, 107 right-handers, age range = 18-65, mean age = 36.18, *SD* = 11.33) completed the experiment. This sample size was based on a small predicted effect for the critical three-way interaction. Participants received $0.90 for their participation, which required approximately 25 minutes.

Materials. We created 32 experimental sentences that described a person in locomotion (e.g., *walked*, *trotted*, *sneaked*) relative to an object, according to the template subject-locomotion verb-preposition-object-optional modifier (e.g., *The doctor walked toward the cabinet where he kept his records*). The subject referred to the character picture participants had just seen and the preposition was always *toward* or *away from*. In addition, we created 32 filler sentences that did not describe locomotion and did not contain the preposition *toward* or *away* (e.g., *The doctor switched on the lights in his office*). These sentences were followed by an unrelated probe object. To make sure participants read for comprehension, 16 filler trials were followed by a question (e.g., *Did the doctor switch the lights off?*). Furthermore, we created eight practice trials.

Contextual pictures were line drawings of eight different characters that were named by their professions (e.g., the doctor, the baker, the hunter). The pictures were

fitted to an area of 200 x 100 pixels, positioned in the center of the screen, and flipped along the vertical axis to create two directions of orientation (i.e., left or right). Probe pictures were line drawings of 64 familiar objects, taken from various online libraries. The probe object was fitted to an area of 200 x 200 pixels whose center was positioned 380 pixels to the left or the right of the center of the screen.

Design. This experiment employed a 2 (character orientation: left vs. right) x 2 (direction: toward vs. away) x 2 (object location: left vs. right) design. All independent variables were manipulated within participants. Eight counterbalanced lists were created, so that each probe object was assigned to all eight conditions across participants. Participants saw each character eight times, but item effects for contextual pictures were not analyzed. Trials were presented in random order.

Procedure. Participants first read instructions, which encouraged them to remove possible distractions by toggling the full-screen mode of their browser window and to respond as quickly as possible by keeping their fingers on the appropriate keys throughout the experiment. Then they performed eight practice trials with feedback (the word *correct* or *incorrect* displayed for 1.5 s after the response to the probe pictures and comprehension questions). A trial started with presentation of the contextual picture for 2 s. Then the sentence was displayed (centered, in black 14-point Arial font against a white background). Participants pressed *P* when they had finished reading the sentence. This triggered a fixation cross for 0.5 s, after which the probe object appeared. Participants were asked to press *L* if the object had been mentioned in the sentence and *A* if it had not been mentioned. Figure 1 displays the event sequence for a given experimental trial.

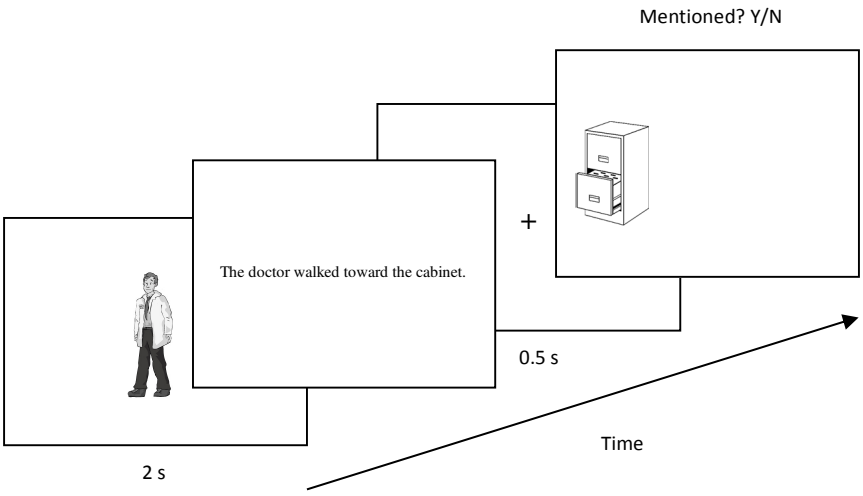


Figure 1. Event sequence in an experimental trial in Experiment 1.

Results

Participants were removed when their reported native language was not English and when they indicated they performed the experiment in a highly noisy and distractive environment (at least 7 on a 9-point scale). This resulted in removal of four participants, who were not replaced. The pre-set inclusion criteria for accuracy across experimental and filler trials (> .80) and comprehension questions (> .60) did not prompt us to remove any additional participants. Next, trials with response times shorter than 0.3 s (because this could hardly reflect perceiving and matching of the probe picture) or longer than 5 s and trials with sentence reading times longer than 10 s (because of the undesirably long delay between the contextual picture and the probe) were excluded. This resulted in removal of 2.7% of the remaining experimental trials. (Analyses were carried out over 3468 trials, 91.8% of the number that was originally collected.)

Table 1
Response Times (ms) and Standard Deviations for All Experiments.

			Exp 1		Exp 2		Exp 3		Exp 4	
Orientation	Direction	Location	RT	*SD*	RT	*SD*	RT	*SD*	RT	*SD*
Left	Away	Left	1358	665	833	329	1028	440	1045	441
		Right*	1302	674	844	342	1077	507	1136	565
	Toward	Left*	1307	574	820	312	1036	459	1013	442
		Right	1273	629	807	332	1069	450	1092	496
Right	Away	Left*	1380	678	827	336	997	429	1016	430
		Right	1332	675	792	290	1074	491	1111	495
	Toward	Left	1323	645	818	326	1020	424	1016	435
		Right*	1291	651	798	316	1072	468	1055	413

Note: asterisks indicate conditions where orientation, direction, and location are congruent.

Table 1 gives the mean response times (computed over correct responses) per condition. The threshold for statistical significance was set at $\alpha = .05$. The three-way interaction Orientation x Direction x Location was not statistically reliable[1], $b = .001$, 95% CI [-.017 .019], Wald(1) = .017, $p = .9$. $r = .032$.[2] The regression coefficient was very

[1] Following Kline (2004) and Cumming (2012), we avoid the term *significant* wherever possible. A small *p*-value says nothing about the size or the importance of an effect, unlike this term is often understood to imply.

[2] The effect size *r* was calculated as follows: $r = \sqrt{\frac{\left(\frac{b}{SE}\right)^2}{\left(\frac{b}{SE}\right)^2 + df}}$ (Rosnow & Rosenthal, 2005).

small, suggesting a negligible influence of this interaction on response times. None of the two-way interactions were statistically reliable, *b*s ≤ .005, *p*s ≥ .600. *r*s ≤ .051. The regression coefficients indicate differences of 10 ms or smaller, suggesting that response times were not influenced by the interplay between any pair of independent variables.

There was no statistically reliable effect of orientation, b = -.010. 95% CI [-.028 .009], Wald(1) = 1.081, p = .300. r = .102, although responses were on average 20 ms faster[3] in trials with left-facing pictures than in trials with right-facing pictures. There was a statistical effect of direction, b = .019, 95% CI [.001 .038], Wald(1) = 4.377, p = .036, r = .203, indicating that responses were on average 39 ms slower after *away*-sentences than after *toward*-sentences. There was a statistical effect of location, b = .022, 95% CI [.004 .040], Wald(1) = 5.522, p = .019, r = .227, indicating that responses to objects on the right were on average 44 ms faster than responses to objects on the left.

Discussion

In a sentence-picture verification task we found no evidence that readers adopt the perspective of a contextual picture in constructing a spatial representation from a linguistic description: there were no statistical interactions between contextual picture orientation, sentence direction, and probe location on response times. However, there was a reaction time advantage for objects on the right half of the screen. Because the appropriate response key for experimental trials was also on the right, this can be interpreted as a Simon effect (Simon, 1969). Given that any of the theoretically relevant effects would implicate an interaction with sentence type, we consider this effect by itself uninteresting. Nevertheless, it may have acted as a confounder, given that the object on the right might be recognized faster due to other reasons. In Experiment 2, we sought to eliminate the presumable Simon effect by requesting a response that requires the use of a single effector: moving the mouse cursor from a central starting point toward the probe object.

Moreover, standard deviations were larger than usual for experiments of this type (i.e., 649 ms on a grand mean of 1321 ms), possibly masking an effect. These may have been due to variation in the stimulus set, part of which we deliberately included to conceal the purpose of the experiment and to make the sentences more interesting. For instance, not all experimental sentences ended with the probe word, and we used many different verbs of locomotion. To make response times more homogeneous, we reduced all these sources of variation in Experiment 2. Also, we replaced some probe objects that had slow or highly variable reaction times. Finally, to increase statistical power, we used a larger sample.

[3] We used effect coding, so the difference between two levels of a given factor is 2**b*.

Experiment 2

Method

Participants. We recruited 160 participants and ended up with 162 (71 males, 145 right-handers, age range 18-67, mean age = 34.54, *SD* = 11.17) who completed the experiment.[4] Participants received $0.80 for their participation, which required approximately 20 minutes.

Materials. The materials were based on those of Experiment 1, with a number of changes. First, we introduced eight new characters, so that participants saw each character only four times. Second, probe objects with slow or highly variable reaction times (e.g., a scoreboard and a trap) were replaced. Third, we removed the optional modifier from the sentences, such that each sentence ended with the probe word, and reduced variation in locomotion verbs, such that the character *walked* in almost all sentences.

Design. The design was identical to Experiment 1.

Procedure. Participants were instructed to respond by moving the mouse cursor (without clicking) over the probe picture as soon as it appeared. To ensure that the mouse cursor was in the same position at the onset of the probe in each trial, participants had to click within a pair of square brackets, displayed 160 pixels below the sentence, to make the probe object appear. Note that the delay between offset of the sentence and onset of the probe was shorter than in Experiment 1 because we removed the fixation cross. This was done to prevent anticipatory mouse movements away from the square. At the end of each trial the brackets remained visible on a blank screen; clicking between them initiated a new trial. Figure 2 illustrates the event sequence for a given experimental trial. The experiment started with 12 practice trials in which participants did not receive feedback.

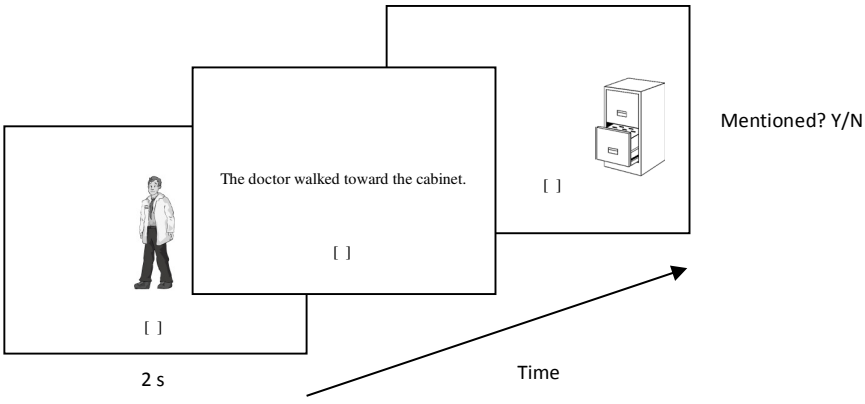


Figure 2. Event sequence in an experimental trial in Experiment 2.

[4] This was due to a technical issue that caused the online portal to not close immediately when the criterion number of participants had begun the experiment. This also occurred in Experiment 4.

Results

We removed the data of 14 participants because they reported a native language other than English, performed the experiment in a noisy environment, or used a touchscreen rather than a mouse or trackpad. In four lists, there was an error in loading the contextual picture in six experimental trials. While this did not disrupt the flow of the experiment, we discarded these trials from analysis. Furthermore, 17 participants failed to meet the accuracy criterion, suggesting they did not perform the task according to the instructions; a common pattern was that participants always pointed to the probe object, regardless of whether it was mentioned in the sentence. Their data were excluded from the dataset. Next, we used the same trial exclusion plan as in Experiment 1 with the exception that there was no cutoff for long response times, because yes-responses were by definition shorter than 3 s. This led to removal of 2.5% of remaining experimental trials. (Analyses were carried out over 3796 experimental trials, 73.2% of the number that was originally collected.)

Response times were defined as the delay between the appearance of the probe and the moment the cursor entered the 200 x 200 area in which the picture was located. Mean response times per condition are given in Table 1. The three-way interaction Orientation x Direction x Location was not statistically reliable, b = -.002, 95% CI [-.011 .006], Wald(1) = .257, p = .610, r = .047. Responses were on average 4 ms faster in congruent trials than in incongruent trials. None of the two-way interactions were statistically significant, $-.006 \leq bs \leq .005$, $ps \geq .190$, $rs \leq .122$. The regression coefficients indicate differences of 12 ms or smaller, suggesting that response times were not influenced by the interplay between any pair of independent variables.

There was no statistical effect of orientation, b = .008, 95% CI [-.001 .016], Wald(1) = 3.044, p = .081, r = .159, although responses were on average 16 ms faster in trials with left-facing pictures than in trials with right-facing pictures. There was no statistical effect of direction, b = .007, 95% CI [-.002 .016], Wald(1) = 2.664, p = .100. r = .149, although responses were on average 14 ms slower after *away*-sentences than after *toward*-sentences There was a statistical effect of location, b = .01. 95% CI [.001 .019], Wald(1) = 5.036, p = .025, r = .205, indicating that responses to probes on the right were on average 20 ms faster than responses to probes on the left.

Discussion

Experiment 2 replicated Experiment 1 in that there was no statistical interaction between picture orientation, sentence direction, and probe location. So, there was no support for the hypothesis that comprehenders use the perspective of a recently seen picture to form a perceptual-like situation model. As in Experiment 1, we did observe an unpredicted response time advantage for objects on the right half of the screen, presumably because it was easier, at least for right-handed individuals, to move the mouse

to that side,[5] which is not interesting with respect to the research question. However, given that the effect is comparable to that in Experiment 1, it becomes more difficult to rule out that objects on the right were verified faster due to other reasons. In Experiment 3, therefore, we tried to rule out this confound by using a go/no-go task with a single response button that was equidistant from both probe locations.

Furthermore, we assumed that one of two unspecified behavioral mechanisms might lead to the predicted difference in reaction times. The first was attentional priming, such that a visual stimulus could be detected faster in a location that matches that of the imagined percept (e.g., Farah, 1985). A reason for not finding the effect might be that the appearance of the probe object served as such a strong exogenous attentional capture that it overrode any potential priming effect. The second possible mechanism was that participants fixated the area of the display where they anticipated the probe object to appear. However, given that it takes approximately 200 ms to plan and execute a saccadic eye movement (Rayner, Slowiaczek, Clifton, & Bertera, 1983), which would be necessary to inspect the object if it appeared in a different location, it seems that an effect should have been detectable under this scenario. In the next experiment, we presented the target probe object along with a distractor object, such that only one object could be perceived foveally at a given time, requiring a saccade from one picture to the other to perceive both. Crucially, we would expect participants' choice of which picture to fixate first to be influenced by the experimental manipulation. If participants fixate the distractor object first, they would have to make a saccade to the other object to determine whether to give a *yes* or *no* response. This transition should take approximately 200 ms – a cost that is easily detectable.

Finally, the contextual picture might not be very effective in conveying left- or rightward motion, because the person seemed to stand still rather than actively walk. This may have prevented participants from integrating the contextual picture with the text. To address this concern, we replaced the pictures of persons with pictures of vehicles that could be more plausibly perceived as moving in the direction of the probe object. While the present multitude of changes does not allow us to isolate the cause of not finding a statistical effect in the previous experiments, we believed these changes were helpful in optimizing our chances of detecting task-induced differences in spatial orienting. It only makes sense to start delineating the conditions under which an effect occurs when there is evidence for it in the first place.

[5] Data from right-handed participants only, then, may give a less biased estimate of the effect of probe location. In an analysis to confirm this post-hoc explanation, the effect of probe location was similar, $b = .009$, 95% CI [.001 .018], Wald(1) = 4.465, $p = .035$, $r = .208$. This suggests that ease of movement or the length of the trajectory may not be the reason for this right-field advantage. Rather, it may derive from participants' visual attention lingering in the right half of the screen after reading the sentence, because it was not re-centered by a fixation cross.

Experiment 3

Method

Participants. We recruited 180 participants, of which 173 (74 males, 145 right-handers, age range 18-8. mean age = 38.43, *SD* = 13.43) completed the experiment. Participants received $0.90 for their participation, which required approximately 25 minutes.

Materials. Contextual pictures were 16 different vehicles (e.g., a car, a van, a tank), fitted to an area of 200 x 200 pixels. Probe pictures were pairs consisting of either a target object and a distractor object or two distractors. Each object was fitted to an area of 200 x 200 pixels whose center was positioned approximately 400 pixels to the left or the right of the center of the screen.

Design. The design was identical to both previous experiments.

Procedure. Pilot testing indicated that self-paced reading with button presses caused a high proportion of 'accidental' presses in no-go trials. Therefore, we presented sentences for a fixed amount of time: 2.5 s. Sentence length was informally judged to be sufficiently uniform to not require variable presentation rates. Participants were instructed to press *B* when one of the pictures that subsequently appeared had been mentioned in the sentence, and to wait for 3 s when neither of the two pictures had been mentioned in the sentence. Figure X illustrates the event sequence for a given experimental trial. In the post-experiment questionnaire, we also asked participants whether clipping of the right picture occurred.

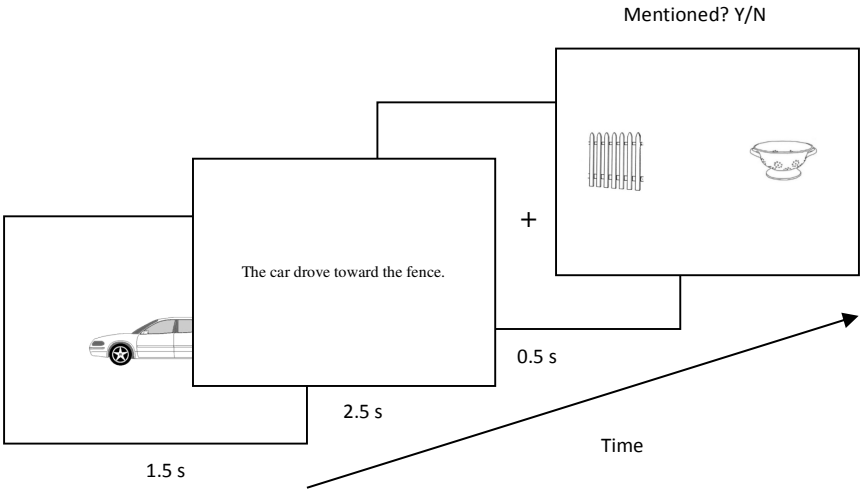


Figure 3. Event sequence in an experimental trial in Experiment 3.

Results

We used the same data exclusion plan as in Experiment 2. Fifteen participants were removed because they reported a native language other than English, performed the experiment in a noisy environment, or experienced problems with clipping of the right picture. We noted very low accuracy (.26) on one experimental item (*pylon*), which was removed. Next, we applied the accuracy criterion of previous experiments. This resulted in removal of another 57 participants. This is an unusually large proportion; however, given that there was a bimodal distribution of participants who performed the task with high and extremely low (zero or close to zero) accuracy, with the proportion of accuracy scores in between being comparable to those of previous experiments, we concluded that the data of the participants who did perform the task accurately were valid and that the only drawback would be a loss of statistical power. Still, the remaining data would allow us to make a fair estimate of the magnitude of the effect of interest, upon which we could decide whether replication using a larger sample size would be worthwhile. The criterion for individual trials (no responses faster than 0.3 sec) did not lead to further exclusion of data.

Mean response times per condition are given in Table 1. The three-way interaction Orientation x Direction x Location was not statistically significant, b = .001, 95% CI [-.012 .014], Wald(1) = .025, p = .870, r = .016. The regression coefficient suggests that responses in congruent trials were a negligible 2 ms faster than responses in incongruent trials. None of the two-way interactions were statistically reliable, -.006 ≤ bs ≤ .006, ps ≥ .37, rs ≤ .089. The regression coefficients indicate differences of 12 ms or smaller, suggesting that response times were not influenced by the interplay between any pair of independent variables.

There was no statistical effect of orientation, b = .006, 95% CI [-.008 .019], Wald(1) = .709, p = .400. r = .083, although responses were on average 12 ms slower in trials with left-facing pictures than in trials with right-facing pictures. There was no statistical effect of direction, b = -.002, 95% CI [-.015 .011], Wald(1) = .112, p = .740, r = .033, although responses were on average 4 ms faster after *away*-sentences than after *toward*-sentences. There was a statistical effect of probe location, b = -.027, 95% CI [-.040 -.014], Wald = 16.427, p < .001, r = .373, indicating that responses to objects on the left were on average 54 ms faster than responses to objects on the right.

Discussion

As in Experiments 1 and 2, the data provided no evidence that participants used the information provided by the contextual picture and the sentence to simulate the location of the probe object. A potential problem was the poor performance by a large proportion of participants; however, the remaining data suggest that the response time difference between congruent and incongruent trials is very small and unlikely to emerge

as statistically reliable in a larger sample. Therefore, we decided to proceed not by replicating this experiment using more explicit instructions, but rather by making two additional changes to increase the task's sensitivity to detect simulation of spatial relations.

First, the contextual pictures were not needed to perform the task correctly and could thus in principle have been ignored. To make sure that participants paid close attention to these pictures, we included questions about visual details (e.g., *Did the van have a license plate?*) in one fifth of the trials. Second, the requirement to respond actively in only half the trials (as was the case in Experiments 2 and 3) might have confused some participants and contributed to their relatively low accuracy. Therefore, we made a button press necessary again: participants were asked to respond with a key on the left if the critical object was shown on the left and with a key on the right if it was shown on the right. This introduced a small difficulty: with only trials in which the object is present, it might be possible for participants to give the appropriate response after inspecting only one object. That is, if the object on the left was not mentioned in the sentence, the object on the right necessarily was, and vice versa. To circumvent this problem, we included a small number of catch trials, where neither object had been mentioned in the sentence. This forced participants to check the second object if the first one was not mentioned in the sentence.

Experiment 4

Method

Participants. We recruited 160 participants and ended up with 162 (87 males, 146 right-handers, age range 18-67, mean age = 35.96, *SD* = 12.15) who completed the experiment. Participants received $0.80 for their participation, which required approximately 20 minutes.

Materials. The materials were based on those of Experiment 3, with a number of changes. First, some object pairs were altered to more carefully balance low-level visual saliency. Darker objects were placed together with equally dark objects, tall shapes with tall shapes, et cetera. Second, the width of the canvas was reduced to 960 px to prevent the object on the right from being clipped on small screens. Third, the objects themselves were reduced to 150 x 150 pixels to make it less likely that both could be accurately perceived at the same time (as might be the case when participants sit at a large distance from the screen). Finally, we reduced the number of filler trials to eight, thereby considerably shortening the duration of the experiment.

Design. The design was identical to all previous experiments.

Procedure. Participants were instructed to press the spacebar when they had read the sentence and to press *A* when the object that was mentioned in the sentence was on the left of the subsequent visual display and *L* when it was on the right. They were also

informed that there were occasional trials in which they would only see two unrelated objects and that waiting 3 s would be the appropriate response. To prevent confusion in these trials, the instructions contained graphical examples of three possible trials and what the appropriate response would be for each.

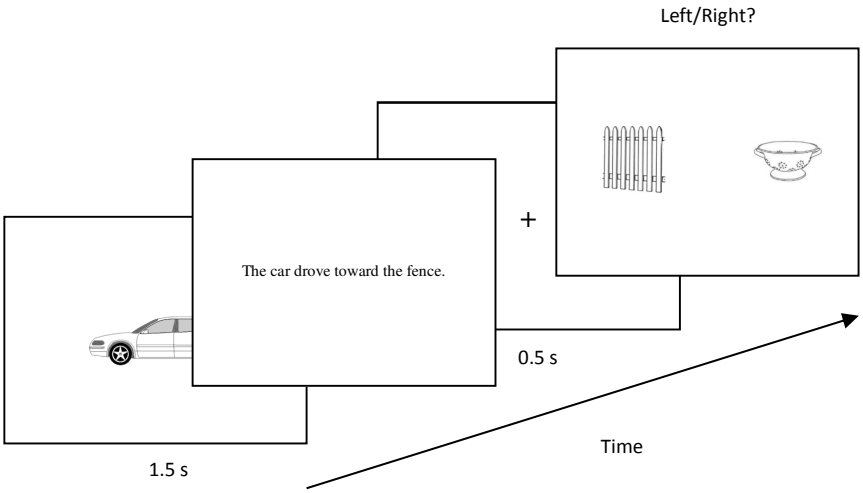


Figure 4. Event sequence in an experimental trial in Experiment 4.

Results

Seven participants were removed because they reported a native language other than English, performed the experiment in a noisy environment, or experienced problems with clipping of the right picture. One item was displayed incorrectly in four lists (i.e., a picture of a boat preceding a sentence about a tractor) and one item was displayed incorrectly in one list. These trials were removed from analysis. Next, three participants were removed due to low accuracy. We used the same data exclusion plan as in Experiment 1, which led to removal of 3.4% of experimental trials. (In total, analyses were carried out over 4591 responses, 82.9% of the number that was originally collected.)

Mean response times per condition are given in Table 1. The three-way interaction Orientation x Direction x Location was not statistically significant, b = .004, 95% CI [-.007 .014], Wald(1) = .409, p = .520, r = .055. The regression coefficient indicates that responses in congruent trials were on average 8 ms faster than responses in incongruent trials. None of the two-way interactions were statistically significant, -.008 ≤ bs ≤ .004, ps ≥ .150, rs ≤ .122. The regression coefficients indicate differences of 16 ms or smaller, suggesting that response times were not influenced by the interplay between any pair of independent variables.

There was no statistical effect of orientation, b = .009, 95% CI [-.001 .020], Wald(1) = 2.834, p = .09, r = 141, although responses were on average 18 ms slower in trials with left-facing pictures than in trials with right-facing pictures. There was a statistical effect of direction, b = .016, 95% CI [.005 .027], Wald(1) = 8.681, p = .003, r = .244, indicating that responses following *toward*-sentences were on average 32 ms faster than responses following *away*-sentences. There was a statistical effect of location, b = -.037, 95% CI [-.048 -.027], Wald(1) = 46.616, $p < .001$, r = .503, indicating that responses to objects on the right were on average 74 ms faster than responses to objects on the left.

Discussion

Like the first three experiments, our fourth experiment provided no statistical evidence that readers were faster to respond to a probe object if it was displayed in a location that was congruent with the contextual picture and the sentence. There was, however, an effect of probe location, such that objects on the left were responded to faster than objects on the right. This most probably indicates that participants preferred to inspect the objects in a left-to-right order. The absence of any interaction effect suggests that they did not change this order as a consequence of processing the contextual picture or the sentence. Also, there was an effect of sentence direction, like in Experiment 1, which we will turn to in the General Discussion.

A somewhat puzzling finding is the discrepancy in absolute response times with Experiment 1, which used a nearly identical procedure. Responses were about 250 ms faster in Experiment 4, even though probe displays in Experiment 4 contained a distractor. A reason might be that to proceed after reading a sentence in Experiment 1, participants pressed the *P* key, which is adjacent to the *L* key, using their right middle finger, whereas they used their thumb and the more distant space bar for this purpose in Experiment 4. Perhaps the slower reaction times in Experiment 1 are due to the relative closeness of these two effectors, which engendered a conflict when responding with the right index finger quickly after responding with the right middle finger.

Meta-Analysis

Four separate experiments failed to find a statistical difference between congruent and incongruent trials. However, the estimate of the effect size was also somewhat inconsistent, as congruent trials were on average faster in Experiments 1, 3, and 4, but slower in Experiment 2, although the confidence intervals of these experiments largely overlapped. A more striking inconsistency appeared to hold for effects regarding sentence direction and probe location, which were as large as 74 ms and statistically reliable in some of the experiments, but not in others, and did not necessarily go in the same direction. To gain a more precise estimate of the effect size, we conducted a meta-analysis (Cumming, 2012). Data from all four experiments were pooled (N = 508) and

regressed on the same set of predictors. We included a by-experiment random intercept and by-experiment random slopes to model the heterogeneity between experiments.

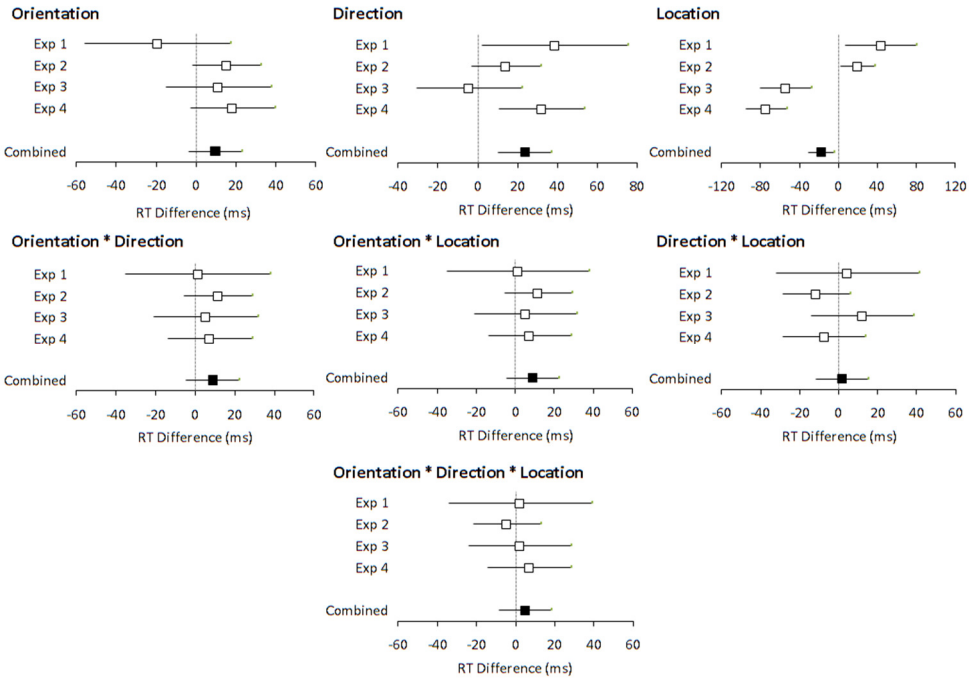


Figure 5. Point estimates and 95% CIs for the effects of all predictors in the individual experiments and meta-analysis. Negative values indicate faster responses, positive values indicate slower responses. Orientation: left (0) vs. right (1); Direction: away (0) vs. toward (1); Location: left (0) vs. right (1).

Figure 5 plots the difference scores in ms and 95% CIs for all predictors across experiments. The three-way interaction Orientation x Direction x Location was not statistically reliable, b = .002, 95% CI [-.004 .009], Wald(1) = .498, p = .480, r = .032. The regression coefficient indicates that responses in congruent trials were on average 4 ms faster than responses in incongruent trials, which is negligible. None of the two-way interactions were statistically significant, -.002 ≤ bs ≤ .004, ps ≥ .200. rs ≤ .058. The regression coefficients indicate differences of 8 ms or smaller, suggesting that response times were not influenced by the interplay between any pair of independent variables.

There was no statistical effect of orientation, b = .005, 95% CI [-.002 .011], Wald(1) = 2.003, p = .16, r = .064, although responses were on average 10 ms slower in trials with left-facing pictures than in trials with right-facing pictures. There was a statistical effect of sentence, b = .012, 95% CI [.005 .018], Wald(1) = 12.123, p < .001, r = .155, indicating that responses following *toward*-sentences were on average 24 ms faster

than responses following *away*-sentences. There was a statistical effect of probe location, b = -.009, 95% CI [-.016 -.002], Wald = 7.015, p = .008, r = .118, indicating that responses to objects on the left were on average 18 ms faster than responses to objects on the right.

To summarize, the meta-analysis suggests that the effect of the three-way interaction is too small to be theoretically meaningful, and not confidently larger than zero. The effect of probe location emerges as statistically reliable, such that objects on the left were responded to faster, but this difference is difficult to interpret, since the presumed causes (i.e., Simon effect, ease of movement, scanning direction) are qualitatively different across experiments. The effect of sentence, however, is interpretable, and suggests a modest but reliable advantage for *toward*-trials.

General Discussion

In four experiments we investigated the effect of contextual pictures on spatial representations of linguistic descriptions. We hypothesized that if comprehenders routinely interpret a sentence with respect to the perspective of a contextual picture, they should be faster to match objects to the text when these appear in congruent locations in external space. None of the experiments provided support for this prediction individually, and a meta-analysis of the combined data of over 500 participants suggests that the effect of congruency between orientation, direction and location on reaction times is close to zero. Before outlining the theoretical consequences of this finding, let us first discuss two methodological issues that may have contributed to the divergence from previous research.

A major departure from earlier experiments is that the visual context, the linguistic stimulus, and the measure of spatial orienting were all temporally separated. Although the blank-screen studies discussed in the introduction (i.e., Johansson et al., 2006; Spivey & Geng, 2001) featured a substantial delay between the presentation of the visual context and the linguistic input, eye movements were measured concurrently with the latter. In our study, visual attention was probed after processing of the sentence had been completed (as indicated by the participants by pressing a button) and a further delay of 0.5 s (except for Experiment 2). However, it should be noted that in the blank-screen studies, eye movements that were presumably language-driven occurred up to 5 s after a specific object had been mentioned, suggesting that it is not the temporal interval alone that accounts for the absence of the predicted effect.

Perhaps more importantly, the modality in which the linguistic stimuli were presented may have sharply limited the degree to which participants could use external space. Reading itself is a process that requires explicit visual attention, interfering with the allocation of visual attention for other purposes. Additionally, the sentences were displayed on the same part of the screen as the contextual picture, rather than above,

below, or beside it. This might have made participants less inclined to use this part of the visual field to 'act out' the described spatial relations. Still, we cannot dismiss the possibility that, precisely because of the inherent burden on visual attention, the modality in which a linguistic message is transmitted leads to qualitative differences in mental representation – a topic that is still little researched.

So, what do these findings tell us about the interaction between visual memory and language processing? They did not support either of our situated interpretation hypotheses, which were predicated on the notion that comprehenders use contextual visual cues and external space to represent the events denoted by linguistic descriptions. Might that mean that spatial relations between characters and objects are encoded in an abstract format, at least during reading? There is one finding, which we did not anticipate at the outset, that goes against this interpretation, namely the response time advantage for trials describing movement toward an object compared to movement away from an object. Another way to see this is that objects located to a person's front (as implied by the verb phrase *walked toward*) were verified faster than objects to a person's back (as implied by the verb phrase *walked away from*).

This result is particularly consistent with the *spatial framework* analysis (e.g., Bryant, Tversky, & Franklin, 1992; Franklin & Tversky, 1990; Franklin, Tversky, & Coon, 1992), which posits that spatial relations in described scenes are conceptualized with respect to the way humans normally perceive and interact with the world. In a series of experiments (Bryant et al., 1992), a cubic array of objects (e.g., a party tent with Halloween decorations) was described from the perspective of a character standing in the middle of that array. When participants were probed with the positions of these objects later (e.g., *pumpkin – above?*), they were fastest to make distinctions along the head-feet axis, which is arguably the predominant axis in spatial cognition (see also Clark, 1973), then along the front-back axis, with front being faster than back, and slowest along the left-right axis. However, when the character was described explicitly as looking into the array from outside, response times changed accordingly, with the difference for front and back disappearing, as the character would be facing both objects. So, it seems that readers have a preference for adopting character-internal viewpoint and encode spatial relations according to where this character is situated.

In keeping with this, the effect of *toward* versus *away* occurred across probe locations and probably has little to do with overt visual attention. Rather, it is attributable to the accessibility of the probe object in the memory representation of the sentence. This accessibility may or may not be determined by the simulation of the object being more or less vivid due to its visibility to the character (Horton & Rapp, 2003; Vandeberg, Eerland, & Zwaan, 2012). The same effect was found for vehicles, at least in Experiment 4. While individuals can plausibly represent the situation as if they were inside a vehicle, there is evidence that they can also assume the perspective of less likely entities, such as a saddle

or a weather vane (Bryant et al., 1992, Experiment 3), suggesting that real-world experience is not a necessary prerequisite.

Interestingly, our raw effect sizes for *toward* versus *away* were considerably smaller than those of Franklin and Tversky (1990), Franklin et al. (1992), and Bryant et al. (1992), where mean response times for items to the back and to the front of a person were more than 100 ms apart. We suggest that task requirements are responsible for this difference. In each of these studies participants gave explicit spatial judgments: the name of an object from a story was briefly displayed and followed by a probe word reflecting one of six directions (e.g., *front*, *above*, or *left*). Moreover, participants knew in advance that they were going to be tested on memory for spatial relations. In our study the manipulation of spatial relations was orthogonal to the required response. Indeed, only 18 participants (mostly from Experiments 1 and 2) suspected that the task they had just performed had something to do with spatial relations. Therefore, our results were less likely to be driven by specific encoding strategies, leading to a more modest effect.

In conclusion, the data suggest that spatial relations are not represented in a wholly abstract, propositional format, but that they have a defined perspective, which is most likely character-internal. Quite surprising about our findings is that readers even seem to adopt this perspective in the context of a picture that could plausibly be related to the subsequent description. So, to return to the question posed in the introduction: presenting a pictorial illustration prior to reading does little to constrain how readers construe entities that were not explicitly depicted. Research with auditorily-presented materials is needed to determine whether the apparent robustness of mental perspective-taking tendencies to contextual visual cues is language-general or specific to reading.

Appendix A. Effect Sizes by Subjects and by Items

Table A1. Effect Sizes and Confidence Intervals Based on Participants or Items as Random Factors.

		Participants				Items			
			95% CI				95% CI		
Predictor	Experiment	Est	*LL*	*UL*	*r*	Est	*LL*	*UL*	*r*
Orientation	Exp 1	-.010	-.028	.009	.102	-.008	-.029	.012	.180
	Exp 2	.008	-.001	.016	.159	.006	-.005	.017	.224
	Exp 3	.006	-.008	.019	.083	.009	-.006	.024	.267
	Exp 4	.009	-.001	.020	.141	.011	-.002	.024	.355
	Combined	.005	-.002	.011	.064	.004	-.003	.012	.102
Direction	Exp 1	.019	.001	.038	.203	.022	.001	.042	.430
	Exp 2	.007	-.002	.016	.149	.007	-.003	.018	.293
	Exp 3	-.002	-.015	.011	.033	-.001	-.016	.014	.021
	Exp 4	.016	.005	.027	.244	.017	.004	.030	.509
	Combined	.012	.005	.018	.155	.012	.004	.019	.283
Location	Exp 1	.022	.004	.040	.227	.021	.000	.041	.415
	Exp 2	.010	.001	.019	.205	.007	-.003	.018	.289
	Exp 3	-.027	-.040	-.014	.373	-.027	-.042	-.012	.631
	Exp 4	-.037	-.048	-.027	.503	-.037	-.050	-.024	.791
	Combined	-.009	-.016	-.002	.118	-.011	-.018	-.003	.263
Orientation	Exp 1	.001	-.018	.019	.007	-.003	-.023	.017	.064
* Direction	Exp 2	.006	-.003	.014	.122	.006	-.005	.016	.239
	Exp 3	.003	-.011	.016	.039	.004	-.011	.018	.111
	Exp 4	.004	-.007	.014	.058	.002	-.011	.015	.059
	Combined	.004	-.002	.011	.058	.002	-.005	.009	.050
Orientation	Exp 1	.002	-.016	.021	.025	.001	-.019	.022	.031
* Location	Exp 2	-.006	-.014	.003	.120	-.007	-.017	.004	.274
	Exp 3	.006	-.007	.019	.089	.008	-.007	.023	.238
	Exp 4	-.004	-.014	.007	.058	-.004	-.017	.009	.134
	Combined	.001	-.006	.007	.011	-.001	-.008	.007	.017
Direction	Exp 1	.005	-.013	.023	.051	.004	-.016	.025	.097
* Location	Exp 2	-.002	-.010	.007	.032	-.001	-.011	.010	.025
	Exp 3	-.006	-.019	.007	.087	-.007	-.022	.008	.198
	Exp 4	-.008	-.018	.003	.122	-.008	-.021	.005	.255
	Combined	-.002	-.008	.005	.021	-.003	-.010	.005	.065
Orientation	Exp 1	.001	-.017	.019	.013	.000	-.020	.020	.002
* Direction	Exp 2	-.002	-.011	.006	.047	-.004	-.015	.006	.172
* Location	Exp 3	.001	-.012	.014	.016	.000	-.015	.015	.006
	Exp 4	.004	-.007	.014	.055	.006	-.007	.019	.198
	Combined	.002	-.004	.009	.032	.001	-.007	.008	.022

Chapter 6

Summary and General Discussion

On Monday morning, the boy was walking to school. He was in the company of another boy, the two boys passing a bag of potato chips back and forth between them. The birthday boy was trying to trick the other boy into telling what he would give in the way of a present. At an intersection, without looking, the birthday boy stepped off the curb, and was promptly knocked down by a car. He fell on his side, his head in the gutter, his legs in the road moving as if he were climbing a wall. The other boy stood holding the potato chips. He was wondering if he should finish the rest or continue on to school.

If you have read the introduction to this thesis, you will be familiar with the above passage. As you may have noticed, however, the ending is different. Rather than crying and dropping the potato chips, the other boy does not seem to care about the accident. Also, there is no mention of what the driver does next. All the same, it is a short story by Raymond Carver (1981). It is titled 'The Bath' and is a different rendering of the same narrative, after scrupulous revisions by his editor Gordon Lish.[1]

But did you also notice that the wording of the prior sentences was different? In fact, not a single sentence is the same as in the introduction. It would not be strange if you failed to notice the difference. It illustrates what we do when we comprehend text: we create and keep a mental representation of the described events, not of the text itself. While some of the exact wording may stay with us, this information typically decays rapidly from memory (Bransford et al., 1972). In Bransford et al.'s seminal paper, when asked to indicate whether they had previously read particular statements, participants frequently responded "yes" to novel statements if they were consistent with the spatial mental model they had constructed. This paper was one of the first to propose a distinction between memory representations for sentences and for 'semantic situations'.

In this thesis, I have investigated what children and adults do to construct such situational representations, and what information is and is not encoded in them. To that end, I have used pictures, both in an 'inside-out' and 'outside-in' fashion. In Chapter 2, response times to pictures demonstrate that children's mental representations of text are in some ways analog to perception, encoding at least the shape and orientation of objects in the referential situation. In Chapter 3, I showed that looks toward pictures in a visual scene can be successfully used to study the temporal dynamics of constructing referential coherence. In Chapter 4, looks toward pictures reveal that adult listeners prioritize constructing coherent situation models over detailed event-internal representations. In Chapter 5, response times to pictures indicated that the direction in which previously seen

[1] The edited version was published first in 'What We Talk About When We Talk About Love' (1981). The original manuscript, which was considerably longer, was published two years later in 'Cathedral' (1983). It can also be found in 'Beginners' (2009), which also contains the unedited versions of the other stories. Because of their generally modest length, these parallel versions may be fine material for the psychological experimentation using authentic texts that some researchers have advocated (e.g., Graesser, Millis, & Zwaan, 1997).

agents faced did not affect where readers mentally simulated other objects in the referential situation.

The next sections provide a more extensive summary of these findings and conclude with some reflections on the role of eye movements in language comprehension and implications for educational research and practice. Before proceeding, however, I wish to make clear that he abundance of pictures in this research should not be taken to imply that the product of comprehension *is* a picture or that somewhere in the brain we should expect to find photographic snapshots as a representational format (see also Kieras, 1978; Pylyshyn, 1973). Perceptual symbols, by definition, afford simulations that are partial and sketchy, rather than detailed and complete (Barsalou, 1999). Accordingly, what I investigated in this thesis is whether the hypothesized mental simulations are sufficiently specific to encode the shape and orientation of entities, and sufficiently permeable to input from other modalities to be constrained by specific elements in the visual environment.

Overview of Main Findings

Chapter 2 was concerned with the question of whether 7- to 12-year-old children mentally simulate the events described in single sentences. We believed this to be a crucial skill in text comprehension, because without a sufficiently detailed understanding at the level of individual events, there is little else to build a coherent model from. Previous studies had shown that adult comprehends routinely simulate the shape and orientation of objects (Stanfield & Zwaan, 2001; Zwaan et al., 2002). In Experiment 1, children listened to sentences that described an object in a specific location. Depending on the location, the implied shape or orientation of the object changed, such as in the pairs *Bob saw the pigeon in the sky / nest* and *Tim saw the screw in the wall / ceiling*. After each sentence, participants saw a picture of the object, its shape or orientation either matching or mismatching that implied by the sentence. Their task was to indicate whether the depicted object was mentioned in the sentence. If children perform perceptual simulations of the sentence content, then we should expect responses to matching pictures to be faster than responses to mismatching pictures. We found that children in all grades except 4 show evidence of perceptual simulation. In Experiment 2, children read the sentences aloud. We used their scores on a standardized measure of reading ability to investigate whether effortful word reading interfered with performing simulations. Because capacity-constrained theories (e.g., Just & Carpenter, 1992; Perfetti & Hart, 2002) state that lower-order processes are prioritized over higher-order processes, we expected poor word-level readers to be prevented from constructing mental simulations. Surprisingly, the results were similar to Experiment 1, and even the non-fluent readers showed a reliable response time advantage for matching pictures.

Across the two experiments, the only clear developmental trend pertained to overall response time: as children got older, they responded faster on average. Grade accounted for unique variance when controlling for motor speed, which suggests that the gain in speed resided in the mental operations that preceded the execution of the response (e.g., recognizing and naming the picture, accessing the memory representation of the sentence, comparing it with the label of the picture, selecting a response). The effect of picture match represented a constant that was similar in magnitude to that found in adult samples. Assuming that a mismatch effect can only occur when a mental representation is sufficiently rich and specific – after all, when it does not have a particular shape or orientation, there is nothing to mismatch – we may infer that mental representations were equally specific across grades. So, by the age of seven, the quality of perceptual simulations has reached a stage beyond which no fundamental changes were observed. Although this does not allow us to draw conclusions about development *before* that age, we may surmise that perceptual simulations play a more prominent role in young children's language comprehension than was previously thought.

Chapter 3 described a comparable sample of children (i.e., 6- to 12-year-olds), but targeted a different level of comprehension, namely that of coherence building in an extended narrative. We asked whether children's performance on a test was related to online processing, which we measured by eye-tracking. Specifically, we wanted to know whether differences in how well children showed a moment-by-moment understanding of whom the text was about factored into their scores on a later test. To this end, they listened to a 7-minute story about four animals and concurrently viewed pictures of these animals, while we recorded their eye movements. Afterwards, they answered 15 literal and inferential comprehension questions. Because answering inferential questions presupposes memory for text details, but memory for text details does not guarantee inferential comprehension, we predicted three profiles: high accuracy on both types of question, high accuracy on literal questions but low accuracy on inferential questions, and low accuracy on both types of question.

Children's scores on the test were modeled with latent class analysis. This is a powerful method for clustering participants on unobserved variables. Rather than the expected three classes, we found two, reflecting good and poor overall accuracy. We nevertheless went on to compare these two classes on their online gaze behavior. We found no difference in the extent to which these classes directed their gaze to one of the pictures if it was referenced by name (e.g., *the rabbit*). This serves as a baseline, indicating that both classes paid attention to the pictures while listening. But we also found no difference in the extent to which the classes directed their gaze to one of the pictures if it was referenced by a pronoun (e.g., *he*); in fact, neither group showed a noticeable shift in attention at all. There was one striking difference, however: whenever a

pronoun came along, the good comprehenders were more likely to have its referent fixated beforehand.

While we did not predict this *a priori*, the results suggest that when proficient comprehenders encounter an anaphoric expression, they already have their attention focused on the appropriate discourse entity. Indeed, this seems a far more efficient mechanism than accessing a memory representation every time a pronoun is heard (see also Givón, 1992). The fact that this is possible in the first place, is because the discourse contains subtle cues as to whether an entity mentioned in the current sentence is likely to be mentioned in the next sentence or not. For example, in English, a given entity is likely to continue to be important in the discourse if it is the sentence subject and is mentioned first. These cues are mainly based on probabilistic patterns of co-occurrence, which are acquired over time with exposure to discourse (see also Arnold et al., 2007). Given that not all children receive the same amount and quality of input, it is not surprising to find substantial individual differences in how well they take advantage of these cues. We propose that a large share of children – approximately two-thirds in our sample – fail to make use of these cues, and that this is an additional reason why anaphor resolution may be challenging to them. The consequence of failing to resolve anaphoric expressions is that children are often confused about who does what do whom. This leads to an impoverished situation model, with fewer referential connections between individual event nodes. This, in turn, caused poorer performance on the comprehension test.

To our knowledge, the study in Chapter 3 was the first to combine the use of a naturalistic text and a genuine online measure. This yielded a substantial improvement in terms of ecological validity compared to previous studies. However, it was hard to determine precisely what properties of the discourse surrounding pronouns were responsible for the differences between good and poor comprehenders. This exemplifies a recurring dilemma in language comprehension research: studying processes in relative isolation gives the researcher control over external variables, but taking the linguistic structure of interest from its context may yield a skewed picture of how it would normally be processed. Indeed, the argument has frequently been made that text comprehension research should proceed by carefully combining both approaches (e.g., Magliano & Graesser, 1991).

Chapter 4 took up an intriguing observation from Chapter 3, namely the occurrence of looks toward the protagonists of the stories despite the fact that their appearances ostensibly mismatched those implied by the text. Specifically, the animals were frequently described as performing actions like running and jumping, but the pictures were static representations of the animals in a resting pose. This calls into question to what extent external representations must overlap with the contents of the story to attract visual attention. We addressed this issue more systematically by contrasting looks toward actions versus entities that were described in stories. Participants

listened to discourse-embedded sentences like *The baker did not hesitate and dove into the water* while concurrently viewing pictures of a diving soldier and a standing baker (and two unrelated distractors). Thus, one picture matched the described action, but not the agent, and one matched the agent, but not the action. By tracking their eye movements to these pictures, we were able to determine which of these mismatches participants were the most tolerant to, or, conversely, what information was most helpful for their comprehension at what moment. Because spoken language unfolds incrementally over time, the mismatch with the visual display would only be apparent during the second half of the sentence. To rule out effects of order-of-mention, we created a parallel version of each sentence in which not the noun but the verb came first (e.g., *Without hesitating dove the brave baker into the water*) – which results in fine grammatical sentences in Dutch.

During the second half of the sentence, we observed a strong prevalence of looks toward the picture that matched the agent (i.e., the standing baker). Participants rarely looked at the picture that matched the action (i.e., the diving soldier). This result affords multiple interpretations. First, agents (or protagonists) form a dimension of coherence, while actions do not. That is, a given sequence of sentences from a text is likely to be about the same character, but not the same action. Participants might prefer to focus on information that is consistent with the narrative at the macro-level over information that only pertains to the micro-level. Second, it may be that the observed gaze behavior reflects the most efficient division of labor between 'offloading' memory demands to the environment and doing mental work: it might be easier to look at a picture of a static character and perform some mental transformations to make it suit the described action, than to look at a picture of an action and imagine a different person performing it. More research is needed to tease out these possibilities. Nevertheless, these results make clear that when we process language in context, different levels of representation may rise to prominence.

In **Chapter 5** we investigated whether readers use a pictorial illustration, presented prior to reading, to ground the spatial relations described in a sentence. As in Chapter 1, we used a sentence-picture verification task. Participants read sentences such as *The doctor walked toward the cabinet where he kept the patient's file*, preceded by a picture of a doctor that was facing left or right from their perspective. Their task was to indicate (by pressing a key) whether the picture of an object that was shown next had been mentioned in the sentence. Crucially, this picture was shown either on the far left or right of the screen, thus matching or mismatching the location implied by the combination of the sentence and the contextual picture. By systematically combining the orientation of the contextual picture (facing left or right), the direction described in the sentence (toward an object or away from it), and the location of the probe picture (left or right on the screen), we could assess to what extent spatial attention was be driven by the integration of linguistic and prior visual information.

The first experiment did not support our hypothesis: the advantage of congruent location was not reliably greater than zero. Even with this three-way interaction broken down into two-way interactions, there was no alignment of linguistic and perceptual cues that led to faster or slower responses. To rule out demand effects, we made some changes in the procedure and materials and ran the experiment again. This time, participants' task was to move the mouse toward the picture if the object had been mentioned in the sentence, and to do nothing if it had not. Again, the critical three-way interaction yielded a negligible effect of cue alignment on response time. We went on to conduct yet another experiment, reasoning that showing two objects might be better than one, because in that case the exogenous attentional capture of a picture appearing 'out of nowhere' would not override potential differences in covert spatial attention. Instead, participants would have to make a choice between the two pictures. We predicted that the order in which they would fixate the pictures would change according to where the prior sentence directed their spatial attention. Again, however, the results provided no support for our hypothesis. In a fourth experiment, we made some minor changes, most importantly that participants were encouraged to remember visual details from the contextual picture. Once more, there were no reliable interactions between contextual picture orientation, sentence direction, and probe location.

A meta-analysis of the four experiments (with a total of more than 500 participants) showed that the contextual picture had a negligibly small, statistically unreliable effect on response time. However, the meta-analysis also revealed a response time advantage for objects that were described as being to the character's *front* compared to the character's *back* that was just large enough to be meaningful. This finding echoes earlier research on *spatial frameworks* (e.g., Bryant, Tversky, & Franklin, 1992; Franklin & Tversky, 1990). When representing spatial relations, readers tend to do so from a character-internal (i.e., first-person) viewpoint. As such, objects to a person's front are recognized faster than objects to a person's back, analogous to how we perceive objects in the real world. We suggest that this difference was due to the accessibility of the probed object in memory. An object that a character is walking *toward*, as opposed to *away from*, could be more salient for two related reasons: because it is visible to the character – the explanation that best aligns with the spatial frameworks analysis and also receives some support from research on occluded objects (Horton & Rapp, 2002; Vandeberg, Eerland, & Zwaan, 2012) – or because it is relevant for future action. Further experiments might adjudicate between these two possibilities. In either case, this mental perspective-taking was impervious to prior visual context, suggesting that the influence of recent visual experiences in language comprehension may be restricted to the representation of those objects to which a direct mapping can be made (Wassenburg & Zwaan, 2010).

Eye Movements during Auditory Comprehension

Chapters 3 and 4 investigated the role of eye movements during spoken language comprehension. Their precise function is an exciting research topic. For one, the visual system might extract useful information from the environment. This is especially the case when language comprehension is *embedded* (Spivey & Richardson, 2008; Zwaan, in press). A chemistry teacher might state "I am now pouring this pink-dyed liquid into the Erlenmeyer" whilst pouring a pink-dyed liquid into an Erlenmeyer. Thus, the words map onto actions and objects that are directly perceivable, obviating the need to retrieve representations from long-term memory. Language comprehension can also be *displaced*. This means that the referential situation is separated in space and time from the communicative situation (Hockett, 1958). Reading the passage about Scotty is an example of this. The smaller the overlap between the referential and the communicative situation, the less obvious the function of eye movements becomes. Yet, individuals also move their eyes in a non-arbitrary way when there is nothing[2] to see (Spivey & Geng, 2001), or when the visual information does not seem very helpful for comprehension, such as when the display shows a static array of characters while the language vividly describes interactions between these characters (Richardson & Dale, 2005). The studies in this thesis were closest to the latter scenario. In Chapter 3, the animals were just sitting there, their appearance in no way mimicking the described changes in the story (e.g., jumping, running). Yet, children kept returning to these locations when the animals were referenced in the story – not overwhelmingly, but more than to other locations. In Chapter 4, a relevant action picture was available, but participants preferred to look at the picture that showed a character that was not performing an action. So, eye movements in our experiments may have been more about indexing the environment than acquiring relevant visual data. That is, the display served as an external memory store, in which each of the four potential referential candidates was associated with an oculomotor coordinate relative to the cues within the scene. In Chapter 3, hearing 'the rabbit', automatically activated these coordinates, causing the eyes to move to that location, regardless of whether the visual information was relevant at that time. Looking at that location may have facilitated retrieval of previously encoded information about the rabbit (see Richardson & Spivey, 2000). By the same token, in Chapter 4, looking at the standing baker might have allowed listeners to more easily map incoming information about the baker onto the appropriate representation in memory.

Implications for Education

One of the goals of this thesis is to inform educational practice. While the results from Chapters 2 and 3 are most directly applicable, Chapters 4 and 5 also suggest some

[2] That is, if a blank computer screen counts as 'nothing'. If not, then total darkness probably does (Johansson, Holsanova, & Holmqvist, 2007).

topics for future classroom research. As I have argued repeatedly throughout this thesis, language comprehension is grounded in perception and action. Therefore, fostering the connection between the words in a text and their external referents might facilitate understanding in readers who do not do so already. This is precisely what imagery techniques that encourage children to form a vivid mental image of what they read (either with or without support or guidance on how to do this) aim at. Indeed, a study by Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004) showed that children had better memory for short narratives and a more accurate understanding of the spatial relations described in them when they acted out the stories by physically manipulating toys and imagining moving them, than when they simply read and re-read the stories. This advantage was later shown to extend to moving toys on a computer screen (Glenberg, Goldberg, & Zhu, 2007) and watching others move toys (Marley, Levin, & Glenberg, 2010). In contrast with this, Chapter 2 shows that 7- to 12-year-old children routinely represent perceptual properties of objects in the referential situation. This suggests that training children of this age to consciously *visualize* the content of sentences is unnecessary. After all, how would detailed instructions to map words in a text onto their external referents help if children already do so? These findings might be reconciled if we take a closer look at the task characteristics. Most importantly, in the studies by Glenberg and colleagues, children had to comprehend *stories* rather than sentences. While the imagery cues were given for each sentence individually, constructing grounded representations might not be as self-evident in the context of a narrative as processing isolated sentences suggests – to which the results of Chapter 4 testify. What manipulation and imagined manipulation instructions might have accomplished was making the process of situation model updating more transparent, for instance by highlighting consistency in terms of actors, objects, and space (see also Rubman & Waters, 2000).

A closely related way to facilitate reading comprehension might be by providing pictorial illustrations. How do these influence mental representations for narrative text? This question was explicitly addressed in Chapter 5, albeit with adults. The results suggest that 'incidental' visual details in pictures of sentence agents do not cause readers to imagine spatial situations differently. So, the influence of pictures seems negligible in this regard. I should make the caveat here that various parameters of the interaction between visual and linguistic input were not systematically explored. For instance, the realism of the pictures (e.g., a line drawing or a photograph), the length of the text (e.g., a single sentence or a short story), and the temporal sequence of presentation (e.g., pictures before, after, or concurrently with the text) might each influence the extent to which visual information impacts mental representations. Changing any of these parameters could, in principle, induce beneficial or detrimental effects of visual cues on text comprehension. These questions remain for future research.

Finally, Chapter 3 demonstrates that failure to resolve anaphoric expressions can lead to poor situation models. What would be an effective way to remediate these difficulties? One solution would be to meet the poor comprehenders' difficulties by rewriting texts to contain fewer anaphoric pronouns and more context-independent references of story characters. However, although this ad-hoc solution might enhance comprehension of specific texts, it would not make the children more skilled comprehenders. A more effective solution might be to train children's awareness of the discourse-level cues that link pronouns to their antecedents, such as subjecthood, order-of-mention, and recency. An intervention that specifically teaches children to resolve anaphoric pronouns exists and is moderately successful (Dommes, Gersten, & Carnine, 1984). Under this approach, children are mainly encouraged to look *back* for the appropriate referent once an anaphoric pronoun is encountered. This approach may be augmented by encouraging children to look *forward*, focusing on the most reliable cues, so that the appropriate referent is more readily available once a pronoun needs to be resolved – just as it is for proficient comprehenders.

Conclusions

In the forty years since Bransford et al.'s paper, knowledge about how humans process and represent text has greatly increased. The present thesis highlights the role of grounded representations, demonstrating their presence in young readers' sentence comprehension. It also shows that recent visual experiences do not permeate every aspect of these grounded representations, and that in the context of a story, readers may 'skip' them in favor of a coherent situation model. It also touches on the role of prediction in language comprehension, and shows how it can make a difference in building coherent situation models.

To put things in perspective: the construction and retrieval of situation models, remarkably sophisticated as it is, may not always be sufficient for successful comprehension, as particular instances of communication failure between speakers and listeners indicate. Therefore, it seems appropriate to allude once more to the 'Scotty' stories, where such failure is a major undercurrent of the action. 'The Bath', the shorter of the two narratives, ends with the birthday boy's mother receiving a strange call after coming home from an exhausting day at the hospital. The conversation is as follows:

> *"Yes!" she said. "Hello!" she said.*
> *"Mrs. Weiss," a man's voice said.*
> *"Yes," she said. "This is Mrs. Weiss. Is it about Scotty?" she said.*
> *"Scotty," the voice said. "It is about Scotty," the voice said. "It has to do with Scotty, yes."*

The mother, frustrated, hangs up, so we learn in 'A Small, Good Thing'. Having forgotten about the birthday cake and sooner expecting news from the hospital, she does not know it is the baker calling, and that he is referring to the birthday cake that has the name 'Scotty' iced on it. The baker, of course, does not know Scotty is in the hospital. Both the baker and the mother rapidly activated a mental network of experienced events – a situation model. Yet, the two fail to align their situation models and as a result, there is no true understanding (Pickering & Garrod, 2004). The conversation also serves as a reminder of how we may come to wrong conclusions about human language comprehension if we only study words and sentences in isolation (Clark, 1997). After all, a boy in a coma and a chocolate cake are two quite different things.

Then again, the fact that we, as readers, could mentally represent this conversation, its contents, and the perspectives of both interlocutors, reflect on it, and possibly be moved by it – with only a collection of lines on paper in front of us – shows the tremendous effect that our text comprehension abilities have on our mental lives.

References

Ackerman, B. P. (1986). Referential and causal coherence in the story comprehension of children and adults. *Journal of Experimental Child Psychology*, *41*, 336-366.

Ackerman, B. P., & McGraw, M. (1991). Constraints on the causal inferences of children and adults in comprehending stories. *Journal of Experimental Child Psychology*, *51*, 364- 394.

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1061-1070.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*, B79-B87.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502-518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*, 55-71.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583-609.

Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes*, *31*, 137-162.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The immediate use of gender information: Eyetracking evidence of the time-course of pronoun resolution. *Cognition*, *76*, B13-B26.

Arnold, J. E., Tanenhaus, M. K., Altmann, R., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science*, *15*, 578-582.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723-767.

Barr, D. J. (2008). Analyzing 'visual world' eye-tracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*, 457-474.

Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, *137*, 201-207.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Barsalou, L. W., Santos, A., Simmons, W., & Wilson, C. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C.

Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 245-283). Oxford: Oxford University Press.

Bartlett, F. C. (1932). *Remembering*. Cambridge, England: Cambridge University Press.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*, 477-487.

Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-193). Hillsdale, NJ: Erlbaum.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*, 733-764.

Beveridge, M. E., & Pickering, M. J. (2013). Perspective taking in language: integrating the spatial and action domains. *Frontiers in Human Neuroscience*, *7:*577.

Bormuth, J. R., Manning, J. C., Carr, J. W., & Pearson, P. D. (1970). Children's comprehension of between- and within-sentence syntactic constructions. *Journal of Educational Psychology*, *61*, 349-357.

Bouwmeester, S., Vermunt, J. K., & Sijtsma, K. (2012). The latent variable approach as applied to transitive reasoning. *Cognitive Development*, 127, 168-180.

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology*, *75*, 189-201.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193-209.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717-726.

Brunyé, T. T., Ditman, T., Mahoney, C. R., Walters, E. K., & Taylor, H. A. (2010). You heard it here first: Readers mentally simulate described sounds. *Acta Psychologica*, 135, 209-215.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, *31*, 74-98.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*, 489-503.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*, 5-26.

Carpenter, P., & Just, M. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Perceptual and Language Processes* (pp. 275-307). New York: Academic Press.

Carver, R. (1981). *What we talk about when we talk about love*. New York: Knopf.

Carver, R. (1983). *Cathedral*. New York: Knopf.

Carver, R. (2009). *Beginners*: *The original version of What we talk about when we talk about love.* London: Cape.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386-404.

Casteel, M. A. (1993). Effects of inference necessity and reading goal on children's inferential generation. *Developmental Psychology*, *29*, 346-357.

Chafe, W. L. (1994). *Discourse, consciousness, and time.* Chicago: University of Chicago Press.

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia*, *37*, 395-402.

Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition*, *28*, 8-17.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*, 128-144.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*, 567-598.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39.

Clark, H. H. (1973). Space, time, semantics, and the child. In T.E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 28-64). New York: Academic Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84-107.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

Deacon, T.W. (1997). *The symbolic species: The coevolution of language and human brain*. London: Penguin

Dijkstra, K., Yaxley, R. H., Madden, C. J., & Zwaan, R. A. (2004). The role of age and perceptual symbols in language comprehension. *Psychology and Aging*, *19*, 352-356.

Dommes, P., Gersten, R., & Carnine, D. (1984). Instructional Procedures for Increasing Skill-Deficient Fourth Graders' Comprehension of Syntactic Structures. *Educational Psychology*, *4*, 155-165.

Dravida, S., Saxe, R., & Bedny, M. (2013). People can understand descriptions of motion without activating visual motion brain regions. *Frontiers in Psychology*, *4:*537.

Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading. *Psychological Science*, *16*, 913-920.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, *114*, 91-103.

Fecica, A. M., & O'Neill, D. K. (2010). A step at a time: Preliterate children's simulation of narrative movement during story comprehension. *Cognition*, 116, 368-381.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11-15.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 182-196.

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition*, *29*, 336-343.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*, 825-850.

Foertsch, J., & Gernsbacher, M. A. (1994). In search of complete comprehension: Getting "minimalists" to work. *Discourse Processes*, *18*, 271-296.

Franklin, N., & Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, *119*, 63-76.

Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, *20*, 507-518.

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek, & R. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544-564). New York: Oxford University Press.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, *17*, 324-363.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. (1996). Coherence cues mapping during comprehension. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships in the production and comprehension of text* (pp. 3-21). Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language*, *27*, 699-717.

Gerrig, R. J. (1993). *Experiencing narrative worlds*. New Haven, CT: Yale University Press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135-176.

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, *30*, 5-55.

Givón, T. (Ed.) (1983). *Topic Continuity in Discourse: A Quantitative Cross-language Study* (Vol. 3). Amsterdam: John Benjamins.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences*, *20*, 41-50.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*, 905-922.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558-565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory & Language*, *43*, 379-401.

Glenberg, A. M., Goldberg, A. B., & Zhu, X. (2009). Improving early reading comprehension using embodied CAI. *Instructional Science*, *39*, 27-39.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology*, *96*, 424-436.

Glenberg, A. M., Meyer, M., Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*, 69-83.

Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, *26*, 651-658.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*, 311-347.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163-189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371-395.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*, 701-721.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local discourse. *Computational Linguistics*, *21*, 203-225.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*, 335-346.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301-307.

Hirschfeld, G., & Zwitserlood, P. (2010). How vision is shaped by language comprehension: Top-down feedback based on low-spatial frequencies. *Brain Research*, *1377*, 78-83.

Hockett, C. F. (1958). A course in modern linguistics. *Language Learning*, *8*, 73-75.

Holmes, B. C. (1985). The effect of four different modes of reading on comprehension. *Reading Research Quarterly*, *20*, 575-585.

Holt, L. E., & Beilock, S. L. (2006). Expertise and its embodiment: examining the impact of sensorimotor skill expertise on the representation of action-related text. *Psychonomic Bulletin & Review*, *13*, 694-701.

Horton, W. S., & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review*, *10*, 104-110.

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403-430.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*, 151-171.

Johansson, R., & Johansson, M. (2014). Look here, eye movements play a functional role in memory retrieval. *Psychological Science*, *25*, 236-242.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*, 1053-1079.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. B. Long, & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187-203). Hillsdale, NJ: Erlbaum.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *98*, 122-149.

Kaiser, E., & Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, *23*, 709-748.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*, 55-80.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133-156.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*, B79-B89.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Ludtke, J. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology*, *60*, 976-990.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. In D. S. McNamara (Ed.) *Reading comprehension strategies: Theories, interventions, and technologies,* (pp. 27-45). Mahwah, NJ: Erlbaum.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin*, *85*, 532-554.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363-394.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133-159.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, *30*, 481-529.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, *35*, 1009-1051.

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*, 23-42.

Kurby, C. A., & Zacks, J. M. (2013). The activation of modality-specific representations during discourse processing. *Brain and Language*, *126*, 338-349.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition*, *112*, 457-461.

Lahey, M. (1988). *Language disorders and language development*. Needham, MA: Macmillan.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processe*s, *25*, 259-284.

Long, D. L., Seely, M. R., & Oppy, B. J. (1999). The strategic nature of less skilled readers' suppression problems. *Discourse Processes*, *27*, 281-302.

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*, 96-104.

Love, J., & McKoon, G. (2011). Rules of engagement: Incomplete and complete pronoun resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 874-887.

Lutz, M. F., & Radvansky, G. A. (1997). The fate of completed goal information in narrative comprehension. *Journal of Memory and Language*, *36*, 293-310.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science*, *14*, 296-301.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35-54.

Madden, C. J., & Dijkstra, K. (2010). Contextual constraints in situation model construction: An investigation of age and reading span. *Aging, Neuropsychology, & Cognition*, *17*, 19-34.

Madden, C. J., & Zwaan, R. A. (2006). Perceptual representation as a mechanism of lexical ambiguity resolution: An investigation of span and processing time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1291-1303.

Magliano J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*, 193-232.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*, 59-70.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*, 587-604.

Mandler, J. M. (2010). The spatial foundations of the conceptual system. *Language and Cognition*, *2*, 21-44.

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*, 173-192.

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2010). What cognitive benefits do dynamic visual representations of a narrative text afford young Native American readers? *Journal of Experimental Education*, *78*, 395-417.

Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly*, *18*, 89-104.

Masur, E. F. (1997). Maternal labeling of novel and familiar objects: Implications for children's development of lexical constraints. *Journal of Child Language*, *24*, 427-439.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition*, *32*, 1389-1400.

Maughan, B. (1995). Long-term outcomes of developmental reading problems. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, *36*, 357-371.

McCutcheon, A. L. (1987). *Latent class analysis*. Quantitative Applications in the Social Sciences Series No. 64. Thousand Oaks, California: Sage Publications.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440-466.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*, 475-494.

Moberly, P. G. C. (1978). *Elementary children's understanding of anaphoric relationships in connected discourse.* (Doctoral dissertation, Northwestern University).

Morrow, D. G. (1985). Prominent characters and events organize narrative understanding. *Journal of Memory and Language*, *24*, 390-404.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*, 665-681.

Myers, J. L., O'Brien, E. J., Albrecht, J. E., & Mason, R. A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 876-886.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. A. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*, 1-9.

Oakhill, J. V. (1982). Constructive processes in skilled and less-skilled comprehenders' memory for sentences, *British Journal of Psychology*, *73*, 13-20.

Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, *54*, 31-39.

Oakhill, J. V., & Yuill, N. (1986). Pronoun resolution in skilled and less-skilled comprehenders: Effects of memory load and inferential complexity. *Language and Speech*, *29*, 25-37.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes*, *18*, 443-468.

Omanson, R. C., Warren, W. M. & Trabasso, T. (1978). Goals, inferential comprehension and recall of stories by children. *Discourse Processes*, *1*, 337-354.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*, 554-566.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Parsons, L. M. (1987). Imagined spatial transformation of one's body. *Journal of Experimental Psychology: General*, *116*, 172-191.

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *Quarterly Journal of Experimental Psychology*, *62*, 1108-1114.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamins.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 1-22.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*, 1-24.

Pyykkönen, P., Matthews, D. & Järvikivi, J. (2010). Three-year olds are sensitive to semantic rominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*, 115-129.

Rall, J., & Harris, P. L. (2000). In Cinderella's slippers? Story comprehension from the protagonist's point of view. *Developmental Psychology*, *36*, 202-208.

Rapp D. N., van den Broek P. W., McMaster, K. L., Kendeou, P., Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: a perspective for research and intervention. *Scientific Studies of Reading*, *11*, 289-312.

Rayner, K., Slowiaczek, M. L., Clifton, C., & Bertera, J. H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 912-922.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, *29*, 1045-1060.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767-780.

Richardson, D. C., & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*, 269-295.

Rosnow, R. L., & Rosenthal, R. (2005). *Beginning behavioural research: A conceptual primer* (5th ed.). Englewood Cliffs, NJ: Pearson/Prentice Hall.

Roy-Charland, A., Saint-Aubin, J., & Evans, M. A. (2007). Eye movements in shared book reading with children from kindergarten to Grade 4. *Reading and Writing*, *20*, 909-931.

Rubman, C. N., & Waters, H. S. (2000). A, B, seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology*, *92*, 503-514.

Sanford, A. J., & Garrod, S. C. (1981). *Understanding written language*. Chichester, UK: John Wiley & Sons.

Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A., & Holopainen, L. (2008). Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *Learning and Instruction*,*18*, 201-210.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime 1.0.* Pittsburgh, PA: Psychological Software Tools.

Schuil, K. D. I., Smits, M., & Zwaan, R. A. (2013). Sentential context modulates the involvement of the motor cortex in action language processing: an fMRI study. *Frontiers in Human Neuroscience*, *7:*100.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417-424.

Shankweiler, D. (1989). How problems of comprehension are related to difficulties in word reading. In D. Shankweiler & I.Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 35-68). Ann Arbor: University of Michigan Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.

Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology*, *81*, 174-17

Smith, N. B. (1965). *American reading instruction.* Newark, DE: International Reading Association.

Song, H., & Fisher, C. (2005). Who's 'she'? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, *52*, 29-57.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, *65*, 235-241.

Spivey, M., & Richardson, D. C. (2008). Language embedded in the environment. In P. Robbins and M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition.* Cambridge, UK: Cambridge University Press.

Spreen, O. (1987). *Learning disabled children growing up. A follow-up into adulthood.* Lisse, The Netherlands: Swets & Zeitlinger.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153-156.

Suitner, C., & Giacomantonio, M. (2012). Seeing the forest from left to right: How construal level affects the spatial agency bias. *Social Psychological and Personality Science*, *3*, 180-185.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.

van den Broek, P. (1997). Discovering the cement of the universe: The development of event comprehension from childhood to adulthood. In P. W. van den Broek, P. J. Bauer, & T. Bourg (Eds.), *Developmental spans in event comprehension and representation: Bridging fictional and actual events* (pp. 321-342). Mahwah, NJ: Lawrence Erlbaum Associates.

van den Broek, P., Lorch, E. P., & Thurlow, R. (1996). Children's and adults' memory for television stories: The role of causal factors, story-grammar categories, and hierarchical level. *Child Development*, *67*, 3010-3028.

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), Sources of coherence in text comprehension (pp. 353-373). Mahwah, NJ: Erlbaum.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah, NJ: Erlbaum.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*, 238-242.

van Straaten, H. (2013). *Je bent super... Jan!* Amsterdam, The Netherlands: Stichting Collectieve Propaganda van het Nederlandse Boek.

Vandeberg, L., Eerland, A., & Zwaan, R. A. (2012). Out of mind, out of sight: Language affects perceptual vividness in memory. *PLoS ONE*, *7*, e36154.

Verhallen, M. J. A. J., & Bus, A. G., (2011). Young second language learners' visual attention to illustrations in storybooks. *Journal of Early Childhood Literacy*, *11*, 480-500.

Verhoeven, L. (1995). *Drie-Minuten-Toets.* Arnhem, The Netherlands: Cito.

Vermunt, J. K., & Magidson, J. (2008). *Latent GOLD 4.5*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2013). *Latent GOLD 5.0.* Belmont, MA: Statistical Innovations Inc.

Wassenburg, S. I., & Zwaan, R. A. (2010). Readers routinely represent implied object rotation: The role of visual experience. *The Quarterly Journal of Experimental Psychology*, *63*, 1665-1670.

Wykes, T. (1981). Inference and children's comprehension of pronouns. *Journal of Experimental Child Psychology*, *32*, 264-278.

Wykes, T. (1983). The role of inferences in children's comprehension of pronouns. *Journal of Experimental Child Psychology*, *35*, 180-193.

Yang, Q., Bucci, M. P., & Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investigative Ophthalmology & Visual Science*, *43*, 2939-2949.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, 105, 229-236.

Yuill, N., & Oakhill, J. V. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology*, *79*, 173-186.

Zabrucky, K., & Ratner, H. H. (1986). Children's comprehension monitoring and recall of inconsistent stories. *Child Development*, *57*, 1401-1418.

Ziegler, F., Mitchell, P., & Currie, G. (2005). How does narrative cue children's perspective taking? *Developmental Psychology*, *41*, 115-123.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. *Psychology of Learning and Motivation*, *44*, 35-62.

Zwaan, R. A. (2008). Experiential traces and mental simulations in language comprehension. In: M. DeVega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning* (pp. 165-180). Oxford: Oxford University Press.

Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology*, *7*, 1142-1150.

Zwaan, R. A. (in press). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences.*

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLoS ONE*, *7*, e51382.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162-185.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*, 1-11.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, 292-297.

Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224-245). New York, NY: Cambridge University Press.

Zwaan, R. A., Madden, C. J., Yaxley, R. H., & Aveyard, M. E. (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, *28*, 611-619.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 386.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168-171.

Zwaan, R. A., Taylor, L. J., & de Boer, M. (2010). Motor resonance as a function of narrative time: Further tests of the linguistic focus hypothesis. *Brain and Language*, *112*, 143-149.

Samenvatting

Begrijpend lezen is een uitermate belangrijke vaardigheid in het onderwijs. De meeste kennisoverdracht vindt plaats door middel van geschreven teksten, of die nu te vinden zijn in een schoolboek, op een website, of in een aantekeningenschrift. Begrijpend lezen is echter ook een vaardigheid die veel kinderen slechts met de nodige moeite leren beheersen. Om het onderwijs op dit gebied zo effectief mogelijk vorm te kunnen geven, is kennis over de processen die ten grondslag liggen aan zowel succesvol als minder succesvol tekstbegrip van essentieel belang (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007).

Ook buiten het onderwijs speelt tekstbegrip een grote rol in onze levens. Het vermogen om uit gesproken en geschreven taal de ideeën van een spreker of schrijver te reconstrueren geldt als een van de meest complexe cognitieve vaardigheden die mensen bezitten. Hoewel er vele niveaus zijn waarop taalverwerking geanalyseerd en bestudeerd kan worden – denk aan klanken, woordbetekenissen en zinsbouw – is tekstbegrip bijzonder omdat het de coördinatie van al die niveaus veronderstelt. Hoe mensen dit precies doen, is een boeiende wetenschappelijke puzzel die nog lang niet opgelost is. De studie van tekstbegrip is daarom niet alleen om praktische, maar ook om fundamentele redenen van grote waarde. In dit proefschrift heb ik enkele facetten van tekstbegrip onderzocht bij kinderen en volwassenen. Om die te illustreren, citeer ik hier de eerste twee zinnen uit het meest recente Kinderboekenweekgeschenk (van Straaten, 2013):

Ergens midden op de grote oceaan dobbert een klein houten vissersbootje waarvan de mast is afgebroken. Vanuit de kajuit klinkt een stem.

Om te karakteriseren wat de lezer moet doen om deze passage te begrijpen, is het handig om een grof onderscheid te maken tussen twee niveaus van betekenis: het microniveau – het combineren van woordbetekenissen, onder de instructies van de syntaxis, tot een mentale representatie van de stand van zaken die in de tekst beschreven wordt – en het macroniveau – het integreren van deze afzonderlijke representaties tot een coherente geheugenstructuur. Tijdens de eerste zin dient de lezer referenten te vinden voor woorden als *oceaan*, *dobbert* en *mast*, in te zien dat het zinsdeel beginnend met *waarvan* betrekking heeft op *een klein houten vissersbootje* en dat de gebeurtenis zich in de tegenwoordige tijd, dus ongeveer in het *nu* afspeelt, waarbij dobberen een activiteit is die langere tijd in beslag neemt. Tijdens de tweede zin dient de lezer de nieuwe informatie te verbinden aan de eerder gelezen informatie, onder de aanname dat het verhaal op het voorgaande voortborduurt. Daarbij dient de lezer de inferentie te maken dat een kajuit een onderdeel van een schip is; het verhaal zoomt dus in op een specifiekere locatie en een specifiekere gebeurtenis.

Ondanks dit voorbeeld vraagt de formulering ‘een mentale representatie van de stand van zaken die in een tekst beschreven wordt’ waarschijnlijk om toelichting. Met een

'mentale representatie' bedoel ik een hernieuwde weergave (*re*-presentatie) van een bepaald concept of een bepaalde relatie, die ergens in het brein tot stand komt. Met 'de stand van zaken die in een tekst beschreven wordt' verwijs ik naar een van de belangrijkste ontdekkingen in het moderne tekstbegriponderzoek, stammend de jaren '70: lezers onthouden niet zozeer de tekst zelf, maar de situatie die door de tekst opgeroepen wordt (Bransford, Barclay, & Franks, 1972) – in dit geval het bemande vissersbootje dat op de oceaan drijft en waar mogelijk iets staat te gebeuren. Een gevolg van de deze definitie is dat betekenis niet vastligt in de tekst, maar door de lezer zelf ingevuld moet worden. De lezer heeft dus een actieve rol in het begripsproces.

Twee hoofdvragen in dit proefschrift zijn: hoe zien die mentale representaties er precies uit en welke processen en mechanismen dragen bij aan de totstandkoming ervan? In mijn poging om deze vragen te beantwoorden heb ik gebruik gemaakt van plaatjes. Ten eerste om te onderzoeken in hoeverre mentale representaties *als plaatjes* zijn: stellen lezers zich bijvoorbeeld voor hoe personen en objecten in een beschreven situatie eruit zien, of zijn mentale afspiegelingen daarvan normaal gesproken[1] meer abstract? Ten tweede kunnen plaatjes dienen om bepaalde aandachtsprocessen te bestuderen. Voorafgaand onderzoek heeft vastgesteld dat waar iemand naar kijkt in belangrijke mate gestuurd wordt door wat hij of zij op dat moment hoort (Cooper, 1974). Met die kennis kunnen we, door middel van eye-tracking, bijvoorbeeld onderzoeken hoe snel bepaalde woorden in een tekst verwerkt worden[2] en of plaatjes die acties afbeelden voor luisteraars een belangrijkere afspiegeling[3] van het verhaal zijn dan plaatjes die personages afbeelden. Ten derde heb ik onderzocht in hoeverre het zien van een plaatje voorafgaand aan het lezen de mentale representatie van de gebeurtenissen in de tekst kan beïnvloeden.

In dit onderzoek heb ik me laten leiden door twee theoretische aannames. Ten eerste: representaties op het microniveau zijn opgebouwd uit *perceptuele symbolen* (Barsalou, 1999). Anders dan in traditionele theorieën over taalbegrip, is onder deze opvatting de betekenis van taaluiting niet vervat in abstracte proposities, maar zijn woorden onlosmakelijk verbonden met patronen van neurale activatie in verschillende modaliteiten. De betekenis van het woord *oceaan*, bijvoorbeeld, kan niet afgelezen worden aan een formele definitie als WATEROPPERVLAK[ZOUT, GROOT], maar wel aan geheugensporen van het zien (vanaf het strand, in een atlas, in films), horen, voelen, etc., van het concept in de wereld. Tijdens het begrijpen van taal worden perceptuele

[1] Ter verduidelijking: het doel van dit onderzoek is om datgene in kaart te brengen wat lezers normaal gesproken doen tijdens het begrijpen van tekst, niet datgene waar ze in ideale omstandigheden toe in staat zijn. Vrijwel elke lezer zal zich desgevraagd een rijk beeld kunnen vormen van situaties die in een tekst beschreven worden. De vraag die ik probeer te beantwoorden is in hoeverre dit soort voorstellingen 'ingebakken' zijn in het leesproces.

[2] Dit wordt over het algemeen uitgedrukt als de tijd tussen het moment dat de spreker begint aan een woord en het moment dat de blik van de luisteraar op het corresponderende plaatje landt.

[3] Dit wordt over het algemeen onderzocht door voor een bepaald tijdvak (bv. de eerste twee seconden na het begin van een woord) de proportie tijd waarin luisteraars naar plaatje X kijken te vergelijken met plaatje Y.

symbolen gecombineerd tot *mentale simulaties*. Die laten zich het beste omschrijven als een heractivatie van de mentale staat die correspondeert met het lichamelijk ervaren (dus zien, horen, voelen, etc.) van de beschreven situatie. Het zijn echter geen volledige en gedetailleerde, maar fragmentarische en schematische afspiegelingen van die ervaringen. Ook zijn ze onbewust. Mentale simulaties vallen dus niet noodzakelijkerwijs samen met de subjectieve beleving van het in beelden voor zich zien van een situatie.

Ten tweede: op het macroniveau staan de begripsprocessen in dienst van het vormen van een *coherente* mentale representatie. Hoewel de aard en de hoeveelheid van inferentiële activiteit afhangt van de capaciteit en de motivatie en doelen van de lezer, ga ik ervan uit dat de fundamentele drijfveer achter het verwerken van tekst een 'zoektocht naar betekenis' is (Bartlett, 1932; Graesser, Singer, & Trabasso, 1994). Bij het creëren van samenhang tussen de binnenkomende informatie en reeds verwerkte informatie houdt de lezer ten minste vijf narratieve dimensies in de gaten: tijd, ruimte, protagonist, intentionaliteit, en causaliteit (Zwaan & Radvansky, 1998). Hierdoor ontstaat een netwerk in het langetermijngeheugen dat doorgaans een *situatiemodel* wordt genoemd.

Dit proefschrift is verder toegespitst op de vraag in hoeverre de micro- en macroniveaus elkaar ondersteunen en met elkaar concurreren om de aandacht van de lezer of luisteraar. Ten slotte besteed ik aandacht aan het ontwikkelingsvraagstuk: kunnen theoretische modellen van tekstbegrip de processen en uitkomsten bij kinderen verklaren? Het tekstbegrip van kinderen en volwassenen verschilt op verschillende punten. Kinderen hebben een kleinere woordenschat (kennen ze het woord *kajuit*?), hebben meer moeite met het automatisch omzetten van geschreven tekst in betekenisvolle representaties (kunnen ze het woord *vissersboot*je vlot lezen of moet het eerst in lettergrepen gehakt worden?) en beschikken niet altijd over de kennis die noodzakelijk is om verbanden te leggen tussen verschillende gebeurtenissen (begrijpen ze dat een schip met een gebroken mast moeilijk kan varen en waarschijnlijk eerder in een storm terecht is gekomen?). In hoeverre leidt dit tot kwalitatief verschillende representaties? Hoofdstuk 2 richt zich op deze vraag.

Hoofdstuk 2: Mentale Simulatie in Jonge Lezers

Hoofdstuk 2 beschrijft een onderzoek naar mentale simulatie in het taalbegrip van kinderen. Een belangrijke theoretische vraag is of simulaties *noodzakelijk* zijn voor tekstbegrip of dat ze veeleer een bijverschijnsel zijn van de mentale bewerking van abstracte symbolen (zie Mahon & Caramazza, 2008). Onderzoek met kinderen kan helpen bij het beantwoorden van die vraag. Er is namelijk reden om aan te nemen dat ze, anders dan volwassenen, geen mentale simulaties zouden vormen als die niet een sterk functionele rol hebben, aangezien ze de oppervlaktestructuur van de taal langzamer verwerken (bv. woordherkenning, syntactische ontleding) en minder vertrouwd zijn met veel van de situaties die in teksten beschreven worden. Bij volwassenen is aangetoond dat

beide beperkingen ervoor zorgen dat mentale simulaties niet vanzelfsprekend plaatsvinden, met een verarmde representatie als gevolg (Holt & Beilock, 2006; Madden & Zwaan, 2006). Los daarvan is het waardevol om in kaart te brengen hoe zinsbegrip bij kinderen zich ontwikkelt en of dit samenhangt met technische leesvaardigheid. Hier kunnen onderwijsinterventies weer op inspelen.

Een populair paradigma in het onderzoek naar mentale simulatie is een verificatietaak waarbij deelnemers zo snel mogelijk moeten aangeven of een afgebeeld object voorkwam in een zojuist gelezen zin. Een voorbeeld is een foto of een tekening van een ei na de zin *Luuk zag het ei in de koekenpan*. Hierbij kan de onderzoeker eigenschappen van het plaatje of de zin manipuleren, zonder dat daarmee het correcte antwoord verandert, bijvoorbeeld door een spiegelei of een intact ei te tonen, of de zin te veranderen in *Luuk zag het ei in de doos*. In al deze gevallen is het antwoord 'ja', hoewel sommige combinaties beter passen dan andere. Als de verschijningsvorm van het getoonde object overeenkomt met de situatie in de zin, spreken we van een *match*, anders van een *mismatch*. Als deelnemers het getoonde object gemiddeld sneller verifiëren in de *match*-conditie dan in de *mismatch*-conditie, dan komt dit, zo is de redenering, doordat tijdens het begrijpen van de zin automatisch bepaalde visuele systemen actief werden.

We namen de verificatietaak af bij kinderen van 7 tot en met 12 jaar (groep 4 tot en met 8). Elk kind kreeg een blok met gesproken zinnen en met geschreven zinnen aangeboden. De resultaten van de luistertaak waren duidelijk: in alle klassen (behalve groep 6) werden plaatjes sneller geverifieerd als deze qua vorm of oriëntatie overeenkwamen met de zin. De tijd die kinderen nodig hadden om te reageren was korter naarmate ze ouder werden; leeftijd was ook een voorspeller wanneer we corrigeerden voor motorische snelheid. In een bepaald stadium van de voorbereiding van de respons waren de oudere kinderen dus sneller dan de jongere kinderen. Alleen, het verschil tussen van *match* en *mismatch* was constant: dit bleef in alle klassen rond de 40 milliseconden. Daaruit kunnen we concluderen dat de *specificiteit* van de mentale simulatie (die bepaalt of er een verschil tussen *match* en *mismatch* optreedt) boven de leeftijd van zeven jaar niet aantoonbaar aan ontwikkeling onderhevig is.

Een vergelijkbaar patroon dook op uit de leestaak, waarvan we de resultaten voor de sterke en de zwakke woordlezers vergeleken: beide groepen waren sneller in de *match*-conditie. Dit was opvallend, aangezien er al enkele decennia bewijs is dat lagere-ordeprocessen geprioriteerd worden tijdens het lezen (Just & Carpenter, 1992) – dus het decoderen van de woorden in de zin krijgt voorrang boven het integreren van woordbetekenissen. Als die lagere-orderprocessen veel aandacht vereisen, gaat dat ten koste van de kwaliteit van de representatie. Deze data suggereren dat mentale simulaties een fundamenteel onderdeel zijn van tekstbegrip. Hoewel dit geen direct en sluitend bewijs is voor de noodzakelijkheid ervan, is het niet waarschijnlijk dat ze bij jonge kinderen optreden als ze geen centrale rol spelen bij het begrijpen van taal.

Hoofdstuk 3: Variatie in Oogbewegingen, Referentiële Coherentie en Tekstbegrip

In Hoofdstuk 3 zetten we een stap van het microniveau naar het macroniveau, in het bijzonder het opbouwen en onderhouden van referentiële coherentie tijdens het luisteren naar een langer verhaal. Hierbij gaat het om de *wie*- en *wat*-vragen die een lezer zou kunnen stellen: over wie of wat gaat deze zin, en wat wordt er over die persoon of die zaak gezegd? Een belangrijke deelvaardigheid is het kunnen begrijpen van anaforische uitdrukkingen: woorden zoals *hij* en *hem*, die terugwijzen naar een concept dat eerder in de tekst genoemd is. Omdat het begrijpen van zulke woorden deels een actief proces is, dat het maken van inferenties vereist, valt er de nodige variatie te verwachten in hoe vaardig kinderen hierin zijn. Kunnen we verschillen hierin meten *tijdens* het begrijpen van tekst en zijn die verschillen te koppelen aan hoe goed kinderen de tekst onthouden? Dit kan waardevolle informatie opleveren, omdat we niet alleen geïnteresseerd zijn in de vraag *of* kinderen voornaamwoorden begrijpen, maar ook *hoe snel* ze dit doen.

Om dit te onderzoeken, gebruikten we de hierboven beschreven eye-trackingmethodologie. Kinderen van 6 tot 12 jaar luisterden een verhaal van 7 minuten over de avonturen van een viertal dieren in het bos (een egel, een konijn, een eekhoorn en een muis). Ondertussen registreerden we hun oogbewegingen naar afbeeldingen van de dieren, die gedurende het hele verhaal te zien waren. Na afloop kregen ze een mondelinge begripstoets die bestond uit vijftien vragen. Met deze data benaderden we de onderzoeksvraag in twee stappen. Eerst werden kinderen aan de hand van hun scores op de begripstoets ingedeeld in een groep sterke en zwakke begrijpers. Vervolgens onderzochten we of deze twee groepen tijdens het luisteren verschillend kijkgedrag lieten zien. We voorspelden dat de verschillen het grootst zouden moeten zijn bij de voornaamwoorden (*hij*, *hem*). Deze veronderstellen namelijk begrip van de voorafgaande context, terwijl zelfstandige naamwoorden (*konijn*, *egel*) ook zonder begrip van de context aan de plaatjes gerelateerd kunnen worden. Behalve dat de kans kleiner zou moeten zijn dat zwakke begrijpers uit zichzelf naar het betreffende plaatje zouden kijken, verwachten we ook dat ze dit over het algemeen langzamer zouden doen.

De resultaten lieten een ander patroon zien dan verwacht: er was geen verschil tussen sterke en zwakke begrijpers in de toename van fixaties op het plaatje waar de spreker naar verwees. Echter, de kans dat sterke begrijpers *vooraf* al naar het betreffende plaatje keken, was wel groter, en dan vooral bij anaforische verwijzingen. Wat betekent dit? Een aannemelijke verklaring is dat de groep sterke lezers de referent van het voornaamwoord vaak al voorspeld had. Dit is mogelijk omdat voornaamwoorden inherent ook voorspelbaar *zijn* (Gordon, Grosz, & Gilliom, 1993). Die voorspelbaarheid is mede terug te voeren op structurele eigenschappen van de tekst: de entiteit die als eerste genoemd is in zin A, heeft een grotere kans om verkort te worden tot voornaamwoord in zin B dan andere entiteiten (Gernsbacher & Hargreaves, 1988). In de korte sequentie *Henk belt Bert. Hij heeft nog het een en ander uit te leggen* zullen de meeste volwassen lezers

om die reden *Hij* in de tweede zin opvatten als terugverwijzend naar *Henk*, hoewel ook *Bert* tot een aannemelijke interpretatie leidt.

Het lijkt erop dat goede begrijpers gevoelig zijn voor dit soort informatie en haar gebruiken in hun verwachtingen over hoe het verhaal verder zal gaan, terwijl de minder goede begrijpers haar ofwel niet opmerken, ofwel negeren, ofwel niet toe kunnen passen. Verder onderzoek zal nodig zijn om deze drie verklaringen systematisch te vergelijken. Overigens vonden we dat de leeftijd van de kinderen niet correleerde met de kans dat ze al van tevoren naar de referent van een voornaamwoord keken. Het lijkt hier dus te gaan om een vaardigheid die niet sterk samenhing met leeftijd, maar wellicht meer met andere individuele kenmerken.

Kortom, dit hoofdstuk laat zien dat eye-tracking waardevolle informatie kan verschaffen over de processen die bijdragen aan een coherente mentale representatie. Wat we zonder deze techniek niet te weten waren gekomen, is dat veel werk bij het leggen van referentiële verbanden al van tevoren gedaan wordt – in ieder geval door sterke begrijpers.

Hoofdstuk 4: De Rol van Mentale Simulaties in het Begrijpen van Verhalen

In Hoofdstuk 4 bestuderen we de wisselwerking tussen de twee niveaus van representatie die ik tot nu besproken hebben: dat van de interne structuur van gebeurtenissen (het microniveau) en de relaties ertussen (het macroniveau). Theorieën over *belichaamd* taalbegrip zijn voornamelijk gebaseerd op studies die gebruik maakten van losse zinnen en woorden – wat natuurlijk geen goede afspiegeling is van hoe mensen doorgaans taal gebruiken. Zijn mentale simulaties ook te generaliseren naar meer natuurlijke, gesitueerde vormen van taalverwerking, zoals tekstbegrip?

Een verhaalcontext kan mentale simulaties op verschillende manieren beïnvloeden. Enerzijds ligt het voor de hand dat ze specifieker worden naarmate de lezer over een grotere hoeveelheid beperkende of uitbreidende informatie over relevante entiteiten de beschikt. Dit geldt vooral voor personages en ruimtes, maar in principe ook voor acties. Neem de zin *Peter liep de kamer uit.* Geïsoleerd leidt deze zin waarschijnlijk niet tot een erg levendige simulatie. De representatie die de lezer vormt zou vermoedelijk specifieker zijn als hij of zij eerder had geleerd dat *Peter* een 12-jarig kind was dat zijn been in het gips had en *de kamer* zijn koude en tochtige slaapkamer was. Anderzijds kan de verhaalcontext ervoor zorgen dat bepaalde aspecten van een gebeurtenis meer op de voorgrond treden dan andere. Als de voorafgaande zinnen de toestand van Peters been beschreven hadden, zou de handeling van het lopen zelf van belang zijn. Als ze daarentegen ruziënde ouders hadden beschreven, zou het lopen minder relevant zijn en slaat de lezer een simulatie daarvan mogelijk over. Een reden daarvoor is dat een sterke focus op individuele gebeurtenissen ertoe kan leiden dat er minder tijd en verwerkingscapaciteit over is voor het onderhouden van coherentie. Vooral tijdens het begrijpen van gesproken taal, waarbij het in veel gevallen niet mogelijk is om even te

pauzeren of terug te spoelen, kan die beperkte capaciteit een rol spelen. Hoe noodzakelijk zijn gedetailleerde representaties van gebeurtenissen tijdens het begrijpen van verhalen?

Om dit te onderzoeken, gebruikten we wederom de eye-trackingmethodologie. Deelnemers aan het onderzoek luisterden naar korte tekstfragmenten, waarbij er tijdens de voorlaatste zin steeds vier plaatjes te zien waren. Een van de plaatjes toonde de handelende persoon uit de voorlaatste zin (bv. een bakker) die geen handeling uitvoerde, en een ander plaatje toonde de handeling uit de voorlaatste zin, die uitgevoerd werd door een persoon die niet in het verhaal voorkwam (bv. een duikende soldaat). De twee overige plaatjes lieten ter afleiding personen zien die niets met de betreffende zin te maken hadden.

Deelnemers hadden een sterke voorkeur voor het plaatje dat de juiste persoon liet zien, ongeacht of dit in de zin voor (*De bakker was dapper en dook meteen in het water*) of na de actie genoemd werd (*Meteen dook de dappere bakker in het water*). We kunnen dit interpreteren als ondersteuning voor het idee dat volwassen taalgebruikers het macroniveau prioriteren boven het microniveau. Een alternatieve verklaring is dat de deelnemers een mentaal transformatieproces toepasten op de plaatjes. Twee opties die ongeveer tot hetzelfde resultaat zouden leiden zijn dat de luisteraar naar het plaatje van de bakker kijkt en zich voorstelt hoe die er bij het uitvoeren van de betreffende actie uit zou zien, of naar de duikende soldaat kijkt en zich voorstelt hoe een andere persoon er in die houding uit zou zien. De eerste optie is dan waarschijnlijk de gemakkelijkste, omdat alle visuele elementen al aanwezig zijn en alleen mentaal geroteerd hoeven worden.

Er valt hier een interessante parallel te trekken met de resultaten uit Hoofdstuk 3. Daar keken kinderen vaak naar het dier waar de spreker op dat moment naar verwees, ook als er een actie (bv. rennen of springen) beschreven werd die niet overeenkwam met de afbeelding. Het zou dus goed kunnen dat het kijken naar de afbeeldingen niet als primaire functie heeft om relevante visuele details uit de omgeving te halen, maar dat de oogbewegingen dienen om geheugenprocessen te ondersteunen (zie Richardson & Spivey, 2000).

Hoofdstuk 5: Het Effect van Illustraties op Ruimtelijke Mentale Representaties

Hoofdstuk 5 gaat over de invloed van illustraties op tekstbegrip. De eerder besproken theorieën laten toe dat recente visuele ervaringen een sterke invloed hebben op mentale simulaties. Eerder onderzoek toonde bijvoorbeeld aan dat deelnemers die een lange reeks afbeeldingen hadden benoemd ongeveer 20 minuten later in een aantoonbaar ongerelateerde leestaak nog steeds gevoelig waren voor de oriëntatie van die afbeeldingen (Wassenburg & Zwaan, 2010). We onderzochten of het zien van een illustratie van een van de personages voorafgaand aan een tekst zou fungeren als ‘anker’ voor andere elementen in de referentiële situatie. Maakt het uit of een personage vanuit het perspectief

van de lezer naar links of naar rechts kijkt bij het representeren *waar* objecten zich bevinden?

Ter beantwoording van deze onderzoeksvraag gebruikten we opnieuw een verificatietaak, zoals beschreven bij Hoofdstuk 2. Het verloop van het experiment was als volgt: in elke trial zagen deelnemers kort een afbeelding van een persoon, die vanuit hun perspectief naar links of rechts leek te lopen. Vervolgens lazen ze een zin waarin die persoon naar een object toe of van een object weg liep. Daarna zagen ze een nieuwe afbeelding, waarvan ze moesten aangeven of hierop het object te zien was dat in de zin beschreven was. Deze afbeelding verscheen uiterst links of uiterst rechts op het scherm. We voorspelden dat deelnemers sneller op het object zouden reageren als de locatie overeenkwam met de combinatie van de ruimtelijke oriëntatie van de persoon en de beschreven richting.

De resultaten boden geen ondersteuning voor de hypothese: in congruente trials reageerden deelnemers even snel als in incongruente trials. In drie vervolgexperimenten veranderden we steeds een klein aspect van de taak waarvan we vermoedden dat het ervoor zou kunnen zorgen dat we het voorspelde effect niet vonden. Zo gebruikten we voertuigen in plaats van personen, omdat die beter in staat zouden kunnen zijn om de suggestie van beweging te wekken, en lieten we na de zin niet een, maar twee objecten zien, om uit te sluiten dat het plotselinge verschijnen van een object zo'n sterke exogene aandachtstrekker was dat subtiele verschillen in visuele aandacht gemaskeerd werden. Echter, in geen enkele van de vervolgexperimenten was er een verschil tussen congruente trials en incongruente trials.

Om het verzamelde bewijs goed op waarde te schatten, voerden we een meta-analyse uit, waarbij alle resultaten samengenomen werden als betrof het één groot experiment. Hieruit bleek dat het gemiddelde verschil tussen congruente en incongruente trials verwaarloosbaar klein en statistisch niet betrouwbaar was. Kortom, we vonden geen ondersteuning voor de hypothese dat 'incidentele' visuele informatie geïntegreerd werd met de talige informatie in de zin. Het is mogelijk dat de leestaak hier niet optimaal geschikt voor was: omdat de expliciete visuele aandacht nodig was voor het lezen van de zin, kon die niet gebruikt worden voor andere doeleinden. Een luistertaak had wellicht andere resultaten opgeleverd. Overigens is het mogelijk dat er juist door die bezetting van visuele aandacht *kwalitatieve* verschillen optreden in de ruimtelijke representaties die mensen vormen tijdens lezen en luisteren – een intrigerend onderwerp voor toekomstig onderzoek. Verder is ook de kanttekening te plaatsen dat de zinnen kort en enigszins triviaal waren. Mogelijk leidde dit ertoe dat deelnemers weinig reden hadden om een gespecificeerde representatie te vormen (zoals besproken bij het voorbeeld *Jan liep de kamer uit* bij Hoofdstuk 4).

Uit de meta-analyse kwam evenwel iets verrassends naar voren: over alle experimenten samengenomen was er statistisch bewijs dat objecten sneller herkend

werden na zinnen die beweging er *naartoe* beschreven dan naar zinnen die beweging er *vandaan* beschreven, ongeacht de oriëntatie van de persoon en de locatie van het object. Dit is consistent met theorieën over ruimtelijke representatie uit de vroege jaren '90 (Franklin & Tversky, 1990): als individuen zich ruimtelijke situaties voorstellen, doen ze dit gewoonlijk vanuit een intern perspectief, waarbij de dimensies boven-onder, voor-achter en links-rechts gerelateerd worden aan het menselijk lichaam. Hoe verklaart dit precies onze data? Als het personage in de zin naar een object toe beweegt, is het object zichtbaar en waarschijnlijk relevant voor een toekomstige handeling. Dit zorgt ervoor dat de lezer het object sterker activeert dan wanneer het personage er van wegloopt; in dat geval is het immers niet zichtbaar of relevant voor een toekomstige handeling. Door die sterkere activatie kon de afbeelding sneller benoemd en met de representatie van de zin vergeleken worden.

Conclusies

Dit onderzoek geeft een beeld van diverse mentale processen die bij kinderen en volwassen plaatsvinden tijdens het begrijpen van tekst. Een overkoepelend thema was het gebruik van plaatjes, op manieren die als 'van-binnen-naar-buiten' en 'van-buiten-naar-binnen' omschreven kunnen worden. Kort samenvattend kan ik op basis van dit proefschrift stellen dat plaatjes aantonen dat de mentale representaties die kinderen maken bij het begrijpen van zinnen in sommige opzichten analoog zijn aan perceptuele ervaringen (Hoofdstuk 2) en daarmee kwalitatief meer lijken op die van volwassenen dan vooraf aangenomen werd. Ook kunnen plaatjes succesvol gebruikt worden om de temporele dynamiek van het onderhouden van referentiële coherentie bloot te leggen in langere verhalen (Hoofdstuk 3) en toont het kijken naar plaatjes aan dat taalgebruikers het opbouwen van een coherent situatiemodel voorrang geven boven het gedetailleerd representeren van specifieke handelingen (Hoofdstuk 4). Ten slotte leidt het bekijken van een plaatje voorafgaand aan het lezen van een tekst er niet toe dat lezers ruimtelijke representaties met betrekking tot dat plaatje vormen (Hoofdstuk 5).

Deze samenvatting begon met de centrale rol van tekstbegrip binnen het onderwijs. Wat kunnen de resultaten in dit proefschrift betekenen voor het onderwijs en toegepast onderwijsonderzoek? Een eerste implicatie betreft het versterken van de verbindingen tussen woorden en perceptuele symbolen. Een van de manieren waarop dit kan is door *visualisatie*, waarbij lezers oefenen met het 'maken van een plaatje in hun hoofd', al dan niet met afnemende ondersteuning van daadwerkelijke afbeeldingen. Aansprekende resultaten zijn gerapporteerd door Glenberg, Gutierrez, Levin, Japuntich en Kaschak (2004). Zij vonden dat een training bestaande uit het 'fysiek naspelen' van verhalen met behulp van speelgoed en vervolgens het 'mentaal naspelen' ervan leidde tot een beter geheugen voor details uit het verhaal en beter begrip van de ruimtelijke relaties die erin beschreven werden, dan het simpelweg lezen en herlezen van de tekst. Die

resultaten generaliseerden later naar situaties waarin kinderen het verhaal naspeelden met virtueel speelgoed op een computer (Glenberg, Goldberg, & Zhu, 2007) en situaties waarin kinderen anderen het speelgoed zagen gebruiken (Marley, Levin, & Glenberg, 2010). De resultaten van Hoofdstuk 2 suggereren echter dat 7- tot 12-jarigen zich routinematig een voorstelling van een zin kunnen maken, zonder dat daar training aan vooraf ging. Wat zou in dat geval de toegevoegde waarde van een interventie kunnen zijn? Een belangrijk verschil is dat het begrijpen van een verhaal, zoals in de studies van Glenberg en collega's, wezenlijk andere eisen aan de taalgebruiker stelt dan het begrijpen van losse zinnen (zie Hoofdstuk 4). Het zou goed kunnen dat visualisaties niet zozeer helpen bij het integreren van woorden, als wel bij het integreren van zinnen, bijvoorbeeld doordat ze de continuïteit van personages, objecten en ruimtes verduidelijken.

Daarnaast suggereren de resultaten dat de aanwezigheid van plaatjes bij tekst de uiteindelijke representatie slechts marginaal beïnvloeden. In ieder geval heeft de oriëntatie van een afbeelding geen ingrijpende gevolgen voor hoe lezers zich een ruimtelijke situatie voorstellen en kunnen luisteraars die naar een plaatje kijken dat niet overeenkomt met de actie die in een verhaal beschreven wordt zich iets later nog wel details over die actie herinneren. Ik moet hierbij opmerken dat de betreffende onderzoeken met volwassen zijn uitgevoerd, die mogelijk relatief minder aandacht besteden aan afbeeldingen dan kinderen – zoals ook kinderen tussen groep 3 en 6 steeds minder op afbeeldingen gericht zijn (Roy-Charland, Saint-Aubin, & Evans, 2010). Ook heb ik slechts zeer beperkt deel uit het scala aan mogelijke relaties tussen tekst en afbeeldingen onderzocht. Het realisme van de afbeeldingen (bv. een lijntekening of een foto), de lengte van de tekst (bv. een enkele zin of een paragraaf) en de volgorde van presentatie (bv. afbeeldingen voor, tegelijkertijd of na de tekst) zijn alle factoren die de wisselwerking tussen linguïstische en visuele informatie kunnen sturen.

Een laatste onderwijsimplicatie betreft het begrijpen van anaforische relaties. In de steekpoef van Hoofdstuk 3 leek ongeveer twee derde van de kinderen hier relatief veel moeite mee te hebben. We zagen ook dat dit samenhing met een verarmde representatie van het verhaal. Eerder onderzoek aangetoond dat een training in het oplossen van voornaamwoorden bijdraagt aan beter tekstbegrip (Dommes, Gersten, & Carnine, 1984). Een nieuw inzicht uit Hoofdstuk 3 is dat goede begrijpers bij het begrijpen van anaforische relaties veel werk al van tevoren doen: door gebruik te maken van structurele eigenschappen van een tekst, is redelijk te voorspellen of een entiteit in de huidige zin als voornaamwoord genoemd zal worden in een volgende zin. Ofschoon het niet zonder meer effectief is om zwakke lezers te leren doen wat sterke lezers doen – immers, dit gaat vaak voorbij aan de oorzaak van het probleem (Kendeou, van den Broek, White, & Lynch, 2005) – is het trainen van vooruitkijkende vaardigheden een interessant onderwerp voor verder onderzoek.

Curriculum Vitae

Jan Engelen was born in Terheijden, The Netherlands, on July 25th, 1985. He completed his secondary education in 2003 at the Stedelijk Gymnasium Breda. He received a Bachelor's degree in Linguistics from the University of Amsterdam in 2006 and a Master's degree (cum laude) in General Linguistics from the same university in 2009, having spent one semester abroad as an exchange student at the University of Iceland. In December 2009, Jan started working as a Ph.D. student at Erasmus University Rotterdam. Besides conducting research on text comprehension, he lost his hair and was involved in teaching courses on statistics, presentation skills, and the psychology of language. Jan is also a co-author of the 'Toolbox', a collection of research-based guidelines for fostering the retention and comprehension of educational material, which is now used in primary schools across Rotterdam.

Peer-Reviewed Publications

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2014). Eye movements reveal differences in children's referential processing during narrative comprehension. *Journal of Experimental Child Psychology*, *118*, 57-77.

Eerland, A., Engelen, J. A. A., & Zwaan, R. A. (2013). The influence of direct and indirect speech on mental representations. *PLoS ONE*, *8*, e65480.

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology, 110*, 659-675.

Submitted Manuscripts

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2014). Does picture orientation constrain spatial situation models? *Manuscript submitted for publication.*

Engelen, J. A. A., Bouwmeester, S., de Bruin, A. B. H., & Zwaan, R. A. (2014). The role of grounded event representations in discourse comprehension. *Manuscript submitted for publication.*

Other Publications

Bouwmeester, S., de Bruin, A. B. H., Camp, G., Engelen, J. A. A., Goossens, N. A. M. C., Tabbers, H. K., & Verkoeijen, P. P. J. L. (2012). *Toolbox: 10 oefenstrategieën uit de geheugenpsychologie voor in de klas.* Rotterdam: Stichting BOOR.

Dankwoord

Voor de totstandkoming van dit proefschrift ben ik vele personen dank verschuldigd, allereerst degenen die direct betrokken zijn geweest bij het onderzoek. Samantha, ik kan er nog steeds niet over uit hoezeer ik het getroffen heb met jou als dagelijks begeleider. Je wist me altijd op het goede moment te prikkelen met kritische vragen of gerust te stellen met een mooi compliment. Bedankt dat je deur altijd open stond voor een praatje, uitleg over een statistisch model of een spontaan onderzoeksidee. Daarnaast was je razendsnelle feedback op manuscripten een ongekende luxe. Anique, ik vond het heel fijn dat je ook vanuit Maastricht zo betrokken bent gebleven bij mijn onderzoek. Je slaagde er als geen ander in om me dingen vanuit een ander perspectief te laten bekijken. Bedankt dat je steeds voor me klaar stond als ik vragen had, over wat dan ook. Rolf, ik ben niet alleen een bewonderaar van je inzichten in taalbegrip en je kijk op wetenschap, maar ook van de manier waarop je mij samen met mijn collega-aio's opgeleid hebt tot kritische en creatieve denkers. Bedankt dat je me zo veel hebt geleerd, op zo'n leuke manier.

Zonder de Stichting Bestuur Openbaar Onderwijs Rotterdam was dit proefschrift er evenmin gekomen. Het was een voorrecht om zo goed gefaciliteerd te worden bij het doen van onderzoek. Ik wil de teams van De Driehoek, De Wilgenstam, Jan Antonie Bijloo, De Plevier, De Klimop en Nelson Mandela bedanken voor hun gastvrijheid en het prettige contact gedurende de afgelopen jaren. Ook de kinderen die hebben deelgenomen aan de experimenten en de ouders die daarvoor hun toestemming hebben gegeven ben ik erkentelijk. Daarnaast gaat mijn dank uit naar de bovenschoolse medewerkers voor hun ondersteuning bij tal van onderzoeksgerelateerde activiteiten.

Graag wil ik ook de leden van de promotiecommissie in dit dankwoord betrekken. De leden van de kleine commissie, prof.dr. Joe Magliano, dr. Katinka Dijkstra en dr. Diane Pecher, bedankt dat jullie de tijd hebben genomen om mijn proefschrift te beoordelen en bereid zijn om hier met mij over te discussiëren. De leden van de grote commissie, prof.dr. Paul van den Broek, prof.dr. Max Louwerse en dr. Huib Tabbers, bedankt dat jullie deel willen nemen aan de oppositie.

Ik zou niet weten hoe ik een onderzoeksproject van vier jaar succesvol af had kunnen sluiten zonder de hulp en het gezelschap van collega's. Speciale dank wil ik betuigen aan de 'club van 1 december', Kim, Martine en Nicole. Het is voor mij heel waardevol geweest om van begin tot eind onze ervaringen te kunnen delen. Voor zowel het vieren van successen als het opvangen van tegenslagen had ik me geen fijnere personen kunnen wensen. Kim en Martine, bedankt voor alle mooie momenten op onze levendige kamer, van bijkletsen op maandagochtend tot muziek op vrijdagmiddag en van het feedback geven op elkaars presentaties tot het bewonderen van elkaars bakkunsten. Nicole, bedankt voor je vrolijkheid en de samenwerking bij alle BOOR-activiteiten. Het was een genoegen om met zo'n goed georganiseerd persoon op te trekken. Mijn latere kamergenootjes, Lisette, Kim O. en Daniel, dank jullie wel voor het warme welkom.

Een bijzonder leerzame en verrijkende ervaring was het schrijven en promoten van de Toolbox. Huib, Peter, Gino, Samantha, Anique en Nicole, bedankt voor de constructieve gesprekken, de gezellige etentjes en de onophoudelijke grappen. Graag wil ik in deze context ook Linda van Tuyl even noemen. Bedankt voor je persoonlijke betrokkenheid en je gedreven aanpak.

In het soms wat ongestructureerde aio-bestaan vond ik in de wekelijkse lab meetings een baken (en een dekmantel voor mijn buitenwetenschappelijke experimenteerzucht met koffie en muffins). Anita, Lisa, Karen, Lysanne, Jacqueline, Wim, Anna, Tulio, Wouter, Nathan en Jim, bedankt voor de onderhoudende onderzoeksbesprekingen, borrels en pubquizavondjes. In het bijzonder wil ik Rolf, Anita en Lisa bedanken voor het organiseren van de waanzinnig leuke schrijfweken. Zitmaaiers, boodschappenkratten en schommels zullen nooit meer hetzelfde zijn. Ook van de C&L-meetings en de O&O-pubgroups heb ik erg veel opgestoken. Tamara, Sofie, Lydia, Gabriela, Noortje, Mario, Gerdien en Vincent, bedankt voor jullie goede ideeën en de vermakelijke discussies.

Tevens is tijdens het schrijven aan dit proefschrift ondersteuning op het technische en administratieve vlak onmisbaar gebleken. De medewerkers van het EBL, Gerrit Jan, Christiaan en Marcel, bedankt voor jullie hulp bij het programmeren van mijn experimenten en het installeren van de eye-tracker op school. De medewerkers van het secretariaat, Mirella, Iris en Anqelique, bedankt voor jullie geduld met al mijn vragen. Mario, Gerdien en Martine, bedankt voor het inspreken van teksten van uiteenlopende lengte. Charly, bedankt voor het maken van de tekeningen voor mijn experimenten en de omslag van dit boek.

Ten slotte ben ik sommige personen ook zonder nadere betrokkenheid bij mijn onderzoek grote dank verschuldigd. Mijn ouders, Ton en Rina: mijn leven als aio begon toen ik, zoekend naar geschikte woonruimte in Rotterdam, als *boomerang kid* een paar maanden bij jullie woonde. Ik ben altijd welkom en daar kan ik jullie niet genoeg voor bedanken. Mijn verdere familie, Karin, Gerard en Marian, bedankt voor jullie steun, belangstelling en hulp bij de vele verhuizingen in Rotterdam. En als laatste, Eveline, dankjewel voor je liefde. Je bent de vrouw van mijn dromen, $p < .001$.