Children's Comprehension of Illustrated Narrative Text:

The Role of Tripartite Representations and Perceptual Simulation

Inaugural-Dissertation
zur Erlangung der Doktorwürde der
Fakultät für Humanwissenschaften
der
Julius-Maximilians-Universität Würzburg

Vorgelegt von
Benedikt Thomas Seger
aus Calw

Würzburg
2020

Erstgutachterin: Prof. Dr. Gerhild Nieding

Zweitgutachter: Prof. Dr. Peter Ohler

Meinen Eltern

Bianka und Andreas Seger

Vorwort

Diese Doktorthesis entstand im Rahmen des Projekts „Die Entwicklung der kognitiven Filmverarbeitung“, gefördert von der Deutschen Forschungsgemeinschaft (DFG; Ni496/9-2).

Mein besonderer Dank gilt zuvorderst meiner Betreuerin, Prof. Dr. Gerhild Nieding, die mir ermöglicht hatte, an diesem Projekt mitzuwirken und mir bei meinem Promotionsvorhaben mit Rat und Tat zur Seite stand. Während des gesamten Betreuungsprozesses fanden ergiebige Fachgespräche statt, die mir sehr dabei geholfen haben, meine Standpunkte zum Forschungsthema immer wieder neu zu reflektieren und die mir immer wieder vorgelebt haben, was wissenschaftliches Arbeiten bedeutet.

Des Weiteren bedanke ich mich bei meinem Zweitgutachter Prof. Dr. Peter Ohler und beim Drittprüfer Prof. Dr. Lynn Huestegge jeweils dafür, dass sie sich bereiterklärt haben, durch die Ausübung dieser Funktionen zum Gelingen meiner Promotion beizutragen.

Mein Dank gilt als nächstes meinen Kolleginnen, insbesondere Dr. Wienke Wannagat und Dr. Juliane Hauf für das Zusammenwirken bei den publizierten Fachartikeln, ohne die meine Promotion in dieser Form nicht möglich gewesen wäre. Darüber hinaus möchte ich mich bei allen aktuellen und ehemaligen Kolleg*innen für die schöne gemeinsame Zeit an der Professur für Entwicklungspsychologie und im Zentrum für Mediendidaktik bedanken.

Vielen Dank auch an alle Leiter*innen, Lehrkräfte, Erzieher*innen und Eltern dafür, dass sie die Durchführung der Untersuchungen an ihren Schulen und Kindergärten ermöglicht haben sowie an alle Kinder und Erwachsene, die an den Studien selbst teilgenommen haben. In dem Kontext gilt mein besonderer Dank den wissenschaftlichen Hilfskräften, Praktikantinnen, Bachelor- und Masterstudentinnen, die an der Erstellung des Stimulusmaterials oder an der Durchführung der Studie beteiligt waren (in alphabetischer Reihenfolge): Ekin Günaydin, Wiebke Herter, Kim Hertinger, Alexander Huber, Anna Laumer, Tania Mousikou und Isabelle Weisenburger.

Zuletzt bedanke ich mich auch bei meiner Familie, die mir während der ganzen Zeit immer den Rückhalt gab, den ich brauchte.

Abstract

This doctoral thesis is part of a research project on the development of the cognitive comprehension of film at Würzburg University that was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft) between 2013 and 2019 and awarded to Gerhild Nieding. That project examined children's comprehension of narrative text and its development in illustrated versus non-illustrated formats. For this purpose, van Dijk and Kintsch's (1983) tripartite model was used, according to which text recipients form text surface and textbase representations and construct a situation model. In particular, predictions referring to the influence of illustrations on these three levels of text representation were derived from the integrated model of text and picture comprehension (ITPC; Schnotz, 2014), which holds that text-picture units are processed on both text-based (descriptive) and picture-based (depictive) paths. Accordingly, illustrations support the construction of a situation model. Moreover, in line with the embodied cognition account (e.g., Barsalou, 1999), it was assumed that the situation model is grounded in perception and action; text recipients mentally simulate the situation addressed in the text through their neural systems related to perception (perceptual simulation) and action (motor resonance). Therefore, the thesis also examines whether perceptual simulation takes place during story reception, whether it improves the comprehension of illustrated stories, and whether motor resonance is related to the comprehension of text accompanied by dynamic illustrations. Finally, predictions concerning the development of comprehending illustrated text were made in line with Springer's (2001) hypotheses according to which younger children, compared with older children and adults, focus more on illustrations during text comprehension (perceptual boundedness) and use illustrations for the development of cognitive skills (perceptual support).

The first research question sought to validate the tripartite model in the context of children's comprehension of narrative text, so *Hypothesis 1* predicted that children yield representations of the text surface, the textbase, and the situation model during text reception. The second research question comprised the assumptions regarding the impact of illustrations on text comprehension. Accordingly, it was expected that illustrations improve the situation model *(Hypothesis 2a)*, especially when they are processed before their corresponding text passages *(Hypothesis 2b)*. Both hypotheses were derived from the ITPC and the assumption that perceptual simulation supports the situation model. It was further predicted that dynamic illustrations evoke more accurate situation models than static ones *(Hypothesis 2c)*; this followed from the assumption that motor resonance supports the situation model. In line with the ITPC, it was

assumed that illustrations impair the textbase *(Hypothesis 2d)*, especially when they are presented after their corresponding text passages *(Hypothesis 2e)*. In accordance with earlier results, it was posited that illustrations have a beneficial effect for the text surface *(Hypothesis 2f)*. The third research question addressed the embodied approach to the situation model. Here, it was assumed that perceptual simulation takes place during text reception *(Hypothesis 3a)* and that it is more pronounced in illustrated than in non-illustrated text *(Hypothesis 3b)*; the latter hypothesis was related to a necessary premise of the assumption that perceptual simulation improves the comprehension of illustrated text. The fourth research question was related to perceptual boundedness and perceptual support and predicted age-related differences; younger children were expected to benefit more from illustrations regarding the situation model *(Hypothesis 4a)* and to simulate vertical object movements in a more pronounced fashion *(Hypothesis 4b)* than older children. In addition, *Hypothesis 4c* held that perceptual simulation is more pronounced in younger children particularly when illustrations are present.

Three experiments were conducted to investigate these hypotheses. Experiment 1 (Seger, Wannagat, & Nieding, submitted).compared the tripartite representations of written text without illustrations, with illustrations presented first, and with illustrations presented after their corresponding sentences. Students between 7 and 13 years old (*N* = 146) took part. Experiment 2 (Seger, Wannagat, & Nieding, 2019) investigated the tripartite representations of auditory text, audiovisual text with static illustrations, and audiovisual text with dynamic illustrations among children in the same age range (*N* = 108). In both experiments, a sentence recognition method similar to that introduced by Schmalhofer and Glavanov (1986) was employed. This method enables the simultaneous measurement of all three text representations. Experiment 3 (Seger, Hauf, & Nieding, 2020) determined the perceptual simulation of vertical object movements during the reception of auditory and audiovisual narrative text among children between 5 and 11 years old and among adults (*N* = 190). For this experiment, a picture verification task based on Stanfield and Zwaan's (2001) paradigm and adapted from Hauf (2016) was used.

The first two experiments confirmed *Hypothesis 1*, indicating that the tripartite model is applicable to the comprehension of auditory and written narrative text among children. A beneficial effect of illustrations to the situation model was observed when they were presented synchronously with auditory text (*Hypotheses 2a*), but not when presented asynchronously with written text (*Hypothesis 2b*), so the ITPC is partly supported on this point. *Hypothesis 2c* was rejected, indicating that motor resonance does not make an additional contribution to the comprehension of narrative text with dynamic illustrations. Regarding the textbase, a general negative effect of illustrations was not observed (*Hypothesis 2d*), but a specific negative effect of illustrations that follow their corresponding text passages was seen (*Hypothesis 2e*); the latter result is also in line with the ITPC. The text surface (*Hypothesis 2f*) appears to benefit from illustrations in auditory but not written text. The results obtained in Experiment 3 suggest that children and adults perceptually simulate vertical object movements (*Hypothesis 3a*), but there appears to be no difference between auditory and audiovisual text (*Hypothesis 3b*), so there is no support for a functional relationship between perceptual simulation and the situation model in illustrated text. *Hypotheses 4a–4c* were investigated in all three experiments and did not receive support in any of them, which indicates that representations of illustrated and non-illustrated narrative text remain stable within the age range examined here.

Zusammenfassung

Die vorliegende Doktorthesis ist Teil eines Forschungsprojektes zur Entwicklung des kognitiven Filmverständnisses an der Universität Würzburg, das von der Deutschen Forschungsgemeinschaft im Zeitraum 2013 – 2019 als Zuwendung an Gerhild Nieding finanziert wurde. In diesem Projekt wurde das Verständnis narrativer Texte mit und ohne Illustrationen bei Kindern sowie dessen Entwicklung untersucht. Zu diesem Zweck wurde van Dijk und Kintschs (1983) Drei-Ebenen-Modell verwendet, demzufolge Textrezipient*innen eine Repräsentation der Textoberfläche und der Textbasis bilden sowie ein Situationsmodell konstruieren. Im Speziellen wurden Vorhersagen in Bezug auf den Einfluss von Illustrationen auf diese drei Textrepräsentationsebenen vom integrierten Modell des Text- und Bildverständnisses (ITPC; Schnotz, 2014) abgeleitet; dieses nimmt an, dass Text-Bild-Einheiten sowohl auf einem textbasierten (deskriptiven) als auch auf einem bildbasierten (depiktiven) Pfad verarbeitet werden. Demzufolge unterstützen Illustrationen den Aufbau eines Situationsmodells. Darüber hinaus wurde mit Bezug auf den Ansatz der verkörperten Kognition (z.B. Barsalou, 1999) angenommen, dass das Situationsmodell im Wahrnehmen und Handeln begründet ist; Textrezipient*innen simulieren demnach die im Text dargestellte Situation durch die neuronalen Systeme, die mit Wahrnehmung (perzeptuelle Simulation) und Handlung (motorische Resonanz) in Verbindung stehen. Deshalb untersucht diese Thesis auch, ob perzeptuelle Simulation während der Textrezeption stattfindet, ob diese das Verstehen illustrierter Geschichten verbessert und ob motorische Resonanz einen Bezug zum Verstehen von Texten mit dynamischen Illustrationen aufweist. Schließlich wurden Vorhersagen bezüglich der Entwicklung des Verständnisses illustrierter Texte anhand von Springers (2001) Hypothesen getroffen, wonach jüngere Kinder während des Textverstehens stärker auf Illustrationen fokussieren als ältere Kinder und Erwachsene (perzeptuelle Gebundenheit) und wonach sie Illustrationen für die Entwicklung kognitiver Fertigkeiten nutzen (perzeptuelle Unterstützung).

Die erste Forschungsfrage zielte darauf ab, das Drei-Ebenen-Modell im Zusammenhang mit dem Verständnis narrativer Texte bei Kindern zu validieren, daher sagte *Hypothese 1* voraus, dass Kinder während der Textrezeption Repräsentationen der Textoberfläche, der Textbasis und des Situationsmodells aufweisen. Die zweite Forschungsfrage umfasste Annahmen bezüglich des Einflusses von Illustrationen auf das Textverständnis. Demnach wurde erwartet, dass Illustrationen das Situationsmodell verbessern (*Hypothese 2a*), vor allem, wenn diese vor den ihr jeweils zugeordneten Textpassagen verarbeitet werden (*Hypothese 2b*). Beide Hypothesen wur-

den hergeleitet aus dem ITPC sowie aus der Annahme, dass perzeptuelle Simulation das Situationsmodell unterstützt. Es wurde ferner vorhergesagt, dass dynamische Illustrationen genauere Situationsmodelle hervorrufen als statische (*Hypothese 2c*); dies folgte aus der Annahme, dass motorische Resonanz das Situationsmodell unterstützt. In Übereinstimmung mit dem ITPC wurde angenommen, dass Illustrationen die Textbasis beeinträchtigen (*Hypothese 2d*), vor allem, wenn diese nach den ihnen zugeordneten Textpassagen präsentiert werden (*Hypothese 2e*). Basierend auf früheren Ergebnissen wurde für die Textoberfläche die Hypothese aufgestellt, dass Illustrationen sich günstig auswirken (*Hypothese 2f*). Die dritte Forschungsfrage nahm Bezug auf den verkörperten Ansatz des Situationsmodells. Hierbei wurde postuliert, dass perzeptuelle Simulationen während der Textrezeption stattfinden (*Hypothese 3a*) und dass diese stärker ausgeprägt sind bei illustriertem im Gegensatz zu nicht-illustriertem Text (*Hypothese 3b*); letztere Hypothese stand in Zusammenhang mit einer notwendigen Voraussetzung der Annahme, dass perzeptuelle Simulation das Verständnis illustrierter Texte erhöht. Die vierte Forschungsfrage stand im Kontext der Annahmen perzeptueller Gebundenheit und perzeptueller Unterstützung und sagte Altersunterschiede voraus; es wurde erwartet, dass jüngere im Gegensatz älteren Kindern in Bezug auf das Situationsmodell mehr von Illustrationen profitieren (*Hypothese 4a*) und vertikale Objektbewegungen stärker simulieren (*Hypothese 4b*). Zudem nahm *Hypothese 4c* an, dass die perzeptuelle Simulation bei jüngeren Kindern vor allem dann stärker ausgeprägt ist, wenn Illustrationen gezeigt werden.

Zur Überprüfung dieser Hypothesen wurden drei Experimente durchgeführt. Experiment 1 (Seger, Wannagat & Nieding, eingereicht) verglich die drei Repräsentationsebenen bei schriftlichem Text ohne Illustrationen, schriftlichem Text mit Illustrationen, die vor dem jeweiligen Text erschienen und schriftlichem Text mit Illustrationen, die danach erschienen. Schüler*innen im Alter von 7 bis 13 Jahren ($N = 146$) nahmen daran teil. Experiment 2 (Seger, Wannagat & Nieding, 2019) erforschte die drei Repräsentationsebenen bei auditivem Text, audiovisuellem Text mit statischen Illustrationen und audiovisuellem Text mit dynamischen Illustrationen in einer Stichprobe von Kindern desselben Alters ($N = 108$). In beiden Experimenten wurde eine Satzrekognitionsmethode ähnlich der von Schmalhofer und Glavanov (1986) angewendet. Diese Methode ermöglicht die simultane Messung aller drei Repräsentationsebenen. Experiment 3 (Seger, Hauf & Nieding, 2020) untersuchte die perzeptuelle Simulation von vertikalen Objektbewegungen bei der Rezeption auditiver und audiovisueller narrativer Texte bei Kindern im Alter von 5 bis 11 Jahren sowie bei Erwachsenen ($N = 190$). Hierbei wurde eine Bildverifikationsaufgabe verwendet, die auf Stanfield und Zwaans (2001) Paradigma aufbaut und von Hauf (2016) adaptiert wurde.

Die ersten beiden Experimente bestätigen *Hypothese 1*, was darauf hindeutet, dass das Drei-Ebenen-Modell auf den Kontext des Verständnisses auditiver und schriftlicher narrativer Texte bei Kindern angewendet werden kann. Eine günstige Auswirkung von Illustrationen auf das Situationsmodell wurde beobachtet, wenn diese synchron mit auditivem Text (*Hypothese 2a*), jedoch nicht wenn diese asynchron mit schriftlichem Text präsentiert wurden (*Hypothese 2b*); dies stellt eine partielle Bestätigung der ITPC in diesem Punkt dar. *Hypothese 2c* wurde verworfen, demnach trägt motorische Resonanz nicht zusätzlich zum Verständnis narrativer Texte mit dynamischen Illustrationen bei. Im Hinblick auf die Textbasis wurde kein genereller negativer Effekt von Illustrationen beobachtet (*Hypothese 2d*), jedoch ein spezifischer negativer Effekt wenn diese der ihnen jeweils zugeordneten Textpassage folgten (*Hypothese 2e*); letzteres Ergebnis steht ebenfalls im Einklang mit der ITPC. Die Textoberfläche (*Hypothese 2f*) scheint von Illustrationen bei auditivem, jedoch nicht bei schriftlichem Text zu profitieren. Die Ergebnisse von Experiment 3 legen nahe, dass Kinder und Erwachsene vertikale Objektbewegungen perzeptuell simulieren (*Hypothese 3a*), es scheint jedoch diesbezüglich keinen Unterschied zwischen auditivem und audiovisuellem Text zu geben (*Hypothese 3b*); folglich wird die Annahme nicht unterstützt, dass perzeptuelle Simulationen beim Aufbau des Situationsmodells bei illustrierten Texten eine funktionale Rolle spielen. Die *Hypothesen 4a–4c* wurden in allen drei Experimenten untersucht und in keinem davon bestätigt; daraus folgt, dass Repräsentationen illustrierter und nicht illustrierter narrativer Texte innerhalb des untersuchten Altersbereichs stabil bleiben.

List of included publications

This doctoral thesis is based on three manuscripts that have been accepted and published in peer-reviewed scientific journals.

Experiment 1 (Appendix A). Seger, B. T., Wannagat, W., & Nieding, G. (2021). Children's surface, textbase, and situation model representations of written and illustrated written narratives. *Reading and Writing, 34*, 1415-1440. doi:10.1007/s11145-020-10118-1

Experiment 2 (Appendix B). Seger, B. T., Wannagat, W., & Nieding, G. (2019). How static and animated pictures contribute to multi-level mental representations of auditory text in seven-, nine-, and eleven-year-old children. *Journal of Cognition and Development, 20*(4), 573—591. doi:10.1080/15248372.2019.1636799

Experiment 3 (Appendix C). Seger, B. T., Hauf, J. E. K., & Nieding, G. (2020). Perceptual simulation of vertical object movement during comprehension of auditory and audiovisual text in children and adults. *Discourse Processes, 57*(5-6), 460-472. doi:10.1080/0163853X.2020.1755801

Contents

1. Introduction

Receiving and understanding text is undeniably a crucial activity in an individual's life in modern human societies. In the past few decades, the emergence of audiovisual (television) and interactive (digital) media have cause people to become more and more exposed to multimedia text including at least two different features: verbal text presented in an auditory or written format and pictorial text that – depending on genre – may include graphic organizers, static or animated illustrations, and video sequences. Children and youth appear to be particularly responsive to new developments in the media world; recent media use surveys have demonstrated this as an ongoing trend (Feierabend, Rathgeb, & Reutter, 2019; Rideout & Robb, 2019). With this in mind, research on children's comprehension of multimedia text is of particular relevance; moreover, it would be valuable to explore how the comprehension of multimedia text develops during childhood.

Early studies in media psychology have compared children's reception and comprehension of television with that of other media, including print (Beentjes & van der Voort, 1991; Meringoff, 1980; Pezdek, Lehrer, & Simon, 1984; Salomon, 1984). However, these early results in the field of multimedia comprehension somehow lack the theoretical underpinnings regarding cognitive processes that underlie media-dependent differences in comprehension and the development of these processes. In the late 1980s, when digital media had begun to appear in many households, the research field of multimedia learning emerged (e.g., Mayer, 1989; Mayer & Anderson, 1992; Mayer & Gallini, 1990; Mayer, Steinhoff, Bower, & Mars, 1995) which has led to the so-called *multimedia principle*, according to which learning from verbal text with pictures is superior to learning from verbal text alone (Mayer, 1997, 2009). Other than earlier research on television comprehension, the multimedia principle and other related principles (see Mayer, 2014 for an overview) has been integrated with prominent theories of cognition such as Baddeley's multicomponent working memory model (e.g., 2002). Although there have been some studies with children (e.g., Rieber, 1990), most investigations in this field have been performed with adult samples so that the implications of this research for the development of text comprehension are tenuous at best.

There is a long research tradition known as *reading comprehension* that deals with text comprehension among children and its development during childhood. A large part of the empirical research in this area has conceived of reading comprehension as an ability or skill and focused on predicting individual differences in reading comprehension by using individual differences in basic cognitive skills such as those related to working memory (e.g., Christopher et al., 2012;

Daneman & Carpenter, 1980; Oakhill, Cain, & Bryant, 2003; Palladino, Cornoldi, De Beni, & Pazzaglia, 2001; Pimperton & Nation, 2010; Seigneuric & Ehrlich, 2005). This kind of research can provide profound insights into the development of text comprehension and help answer questions about how the development of text comprehension is intertwined with the development of related cognitive skills, especially when a longitudinal study design is used. However, as its name implies, reading comprehension research is basically limited to written text. Even though auditory and written text comprehension skills appear to be highly interrelated (Gernsbacher, Varner, & Faust, 1990) and similar as to the cognitive skills that predict them (Kim, 2015; Verhoeven & van Leeuwe, 2008), this research area has not provided much evidence about how verbal and pictorial elements of text each contribute to children's comprehension of multimedia text.

The empirical research addressed in this doctoral thesis was designed to shed more light on the development of multimedia text comprehension during childhood. In doing so, it attempted to combine the strengths of the three research traditions outlined above. First, as in early media comparison studies, the presence or absence of pictorial text elements (*illustrations*) was varied experimentally so that a direct causal relationship between illustrations and comprehension-related behavior measures could be established. Second, like research on multimedia learning, the present work was based on current theories of text comprehension that might be extended in line with the results reported here. Third, this research was sensitive toward development by studying children from a broad age range, which has been the approach in most previous investigations of reading comprehension among children. The experiments reported in this thesis compared illustrated versus non-illustrated versions of text with regard to comprehension-related behavior measures. The combination of these three strengths – the experimental variation of text formats, the foundation of the dependent variables in well-established theories of text processing, and the broad age range of the participating children – make up the innovative character of the present research in the field of text comprehension.

The experiments reported here worked with narrative text (*stories*) referring to situations that are close to everyday life events of children in Western societies. The main reason for this was that, other than expository text usually involved in multimedia learning experiments, events from everyday life are largely unrelated to domain-specific knowledge that should otherwise be accounted for when studying text comprehension (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996; Recht & Leslie, 1988; Schneider, Körkel, & Weinert, 1989).

The *Theoretical Background* chapter begins with a more detailed review of the reading comprehension literature as a starting point for the rationale of this work. Section 2.1 introduces the *tripartite model of text representations* (van Dijk & Kintsch, 1983), which holds that representations of the text surface, textbase, and situation model are formed during text reception. The tripartite model underpins the present research; for that reason, I also discuss how relevant components of the development of reading comprehension (Oakhill, 2020; Verhoeven & Perfetti, 2008) can be subsumed under this more parsimonious and flexible account. Section 2.2 describes the *sentence recognition method* (Schmalhofer & Glavanov, 1986) as an established paradigm for simultaneously assessing the tripartite representations; this method was also used in two of the current experiments. The first research question is related to both the model and the method by asking whether children form text surface and textbase representations and construct a situation model.

Section 2.3 deals with the *comprehension of text and pictures.* Its central element is the introduction of the integrated model of text and picture comprehension (ITPC; Schnotz, 2014) which combines insights from multimedia learning research (e.g., Mayer, 1997) with the tripartite model. The ITPC holds that in text-picture units, the situation model can be obtained on text-based (descriptive) and on picture-based (depictive) paths. It is used as the main reference for the second research question presented here; namely, how illustrations affect children's text surface, textbase, and situation model representations of auditory and written text. Further issues related to text presentation formats are addressed there. It is asked (a) whether illustrations have different impacts on auditory versus written text comprehension (Low & Sweller, 2014); (b) whether it makes a difference in written text when the illustration or the verbal text is presented first (Eitel & Scheiter, 2015); and (c) whether static and dynamic illustrations affect the comprehension of auditory text in different ways (Höffler & Leutner, 2007).

To investigate the impact of illustrations on text comprehension more deeply, Section 2.4 refers to theories addressing the *grounding of situation models in perception and action* (Glenberg & Robertson, 2000; Zwaan & Radvansky, 1998). Accordingly, the situation model can be traced back to a neural simulation of the situation addressed in the text within the recipient's sensory and motor systems. In the first two subsections, empirical evidence for the sensory part of this proposal (i.e., perceptual simulation; see Stanfield & Zwaan, 2001) is outlined based on studies with adults and children; the picture verification method used in one of the experiments is also outlined there. Subsection 2.4.3 refers to the motor part of this assumption (i.e., motor resonance; Glenberg & Kaschak, 2002). In the final part of this section, I discuss how perceptual simulation and motor resonance may contribute to text comprehension. Within the scope of the second research question, the issue is raised whether perceptual simulation and motor resonance can account for the effect of illustrations on the situation model. The third research question is also related to this reasoning; it asks whether perceptual simulation is present during text reception and whether the presence of illustrations has an influence on perceptual simulation.

Finally, Section 2.5 addresses the potential implications of this work for the development of comprehending illustrated text: according to Springer (2001), younger children rely more on perceptual features during text reception than older ones (perceptual boundedness) and use these features to improve their comprehension-related cognitive skills (perceptual support). Therefore, the fourth research question of this thesis addresses the moderating role of age on the impact of illustrations on the situation model, the markedness of perceptual simulation during text reception, and the markedness of perceptual simulation during the reception of auditory versus audiovisual text.

In the *Rationale and Research Methodology* chapter, these four research questions are outlined in detail and the hypotheses are derived. The methodologies of the sentence recognition task used in Experiments 1 and 2 and the picture verification task employed in Experiment 3 are briefly described in the context of data collection and analysis.

Then, a short description of the three *experiments* is given and their results are discussed. Experiment 1 refers to a submitted manuscript (Seger et al.) appearing in Appendix A. There, the representations of written text without illustrations, written text with the text passages presented before their corresponding illustrations, and written text with the illustrations presented first were compared in a sample of children between 7 and 13 years old (in grades 2 through 6). These three text formats had no significant impact on the situation model or the text surface, but the textbase was found to be negatively affected by illustrations when they were presented after their corresponding text passages.

Experiment 2 is discussed in the manuscript in Appendix B (Seger et al., 2019). It investigated the representations of auditory text, audiovisual text with static illustrations, and audiovisual text with dynamic illustrations among children aged 8, 10, and 12 years (grades 2, 4, and 6). Unlike the Experiment 1 results for written text, illustrations were found to improve situation models and text surface representations based on auditory text, whereas the textbase was not affected by illustrations. There were also no significant differences between static and dynamic illustrations regarding any levels of representation. Based on these results, the assumption that perceptual simulation is functionally related to the situation model in illustrated text can be maintained, at least for auditory text.

Therefore, Experiment 3 examined the perceptual simulation of vertical object movements described in a story that is presented in either an auditory or an audiovisual format. Children aged 6 (last preschool year), 8 (grade 2), and 10 (grade 4) and adult university students participated in this experiment. Appendix C contains the manuscript describing Experiment 3 (Seger et al., 2020). The results indicate that children and adults perceptually simulate vertical object movements while they listen to narrative text, but there was no difference between auditory and audiovisual text versions in this regard. Age did not moderate the influence of illustrations on tripartite representations or perceptual simulation in either experiment.

The *General Discussion* chapter integrates the findings from the three experiments with the theoretical and empirical work discussed previously. First, it is shown that children build text surface, textbase, and situation model representations when they process narrative text. Second, the results largely comply with the assumption that children process illustrated narrative text on two paths, one descriptive and one depictive; therefore, it is argued that the ITPC applies to the domain of narrative text. Third, it remains unclear how embodied cognition processes contribute to the situation model in illustrated text. Fourth, there is also no evidence that would favor the perceptual boundedness/perceptual support hypotheses. The methodological limitations of the present research and implications for future studies and educational practice are also discussed.

This thesis is part of a research project that was funded by Deutsche Forschungsgemeinschaft (German Research Foundation, grant no. Ni496/9-2, awarded to Gerhild Nieding).

2. Theoretical Background

The experimental investigations described in this doctoral thesis follow two main objectives. First, they aim to answer the questions of *how children of early school age understand text* and *how text comprehension develops during childhood.* Second, they are conceptualized to gauge the *influence of illustrations on children's comprehension* of written and auditory text.

In her recent review, Oakhill (2020) declares that, owing to the complexity of text comprehension, it is necessary to give attention to its underlying processes when trying to obtain insights into its development, including implications for learning and teaching. Among others, these processes include word decoding, meaning comprehension at word and sentence level, integration of information across sentences, inference making, and, finally, constructing a coherent mental model of the text as a whole (cf. Johnson-Laird, 1983). Oakhill (2020) further points out that although there are ample theoretical frameworks of text comprehension, none focuses primarily on its development. Most research on the development of text comprehension has been dedicated to reading comprehension, and many studies have focused primarily on individual differences regarding comprehension-related abilities or skills (e.g., Johnston, Barnes, & Desrochers, 2008; Oakhill et al., 2003; Tilstra, McMaster, van den Broek, Kendeou, & Rapp, 2009; Verhoeven & van Leeuwe, 2008). However, there is evidence that reading and listening comprehension skills are closely intertwined in the course of development (e.g., Gernsbacher et al., 1990; Kim, 2015; Perfetti, Landi, & Oakhill, 2005; Verhoeven & van Leeuwe, 2008); therefore, it is reasonable to investigate the development of *text comprehension* as a complete construct across different text presentation formats. One motivation for the research outlined here is to contribute to a deeper understanding of the cognitive processes underlying text comprehension during childhood.

But how can all these cognitive processes that are considered to constitute text comprehension be integrated into a cohesive – but still parsimonious – theoretical framework? Van Dijk and Kintsch (1983) proposed the so-called *tripartite model*, according to which three levels of text representation are formed during reading or listening: a text surface, a textbase, and a situation model. The tripartite model was used in the present study of text comprehension among children for five reasons. First, it can explain what text comprehension means; namely, the construction of an appropriate situation model. Second, the tripartite model has been employed successfully in a large body of experiments with adults (e.g., Kintsch, 1998). Third, the model may be adequate to the task of integrating the comprehension-related processes postulated in notable approaches to the development of reading comprehension.

The fourth reason for using the tripartite model in this thesis is the existence of a measurement paradigm that allows for quantification of all three representation levels in a single task, whereby the tripartite model has become a useful framework for experimental research on text comprehension. As noted above, a considerable portion of the research on children's text comprehension to date has dealt with quasi-experimental and correlational data based on individual differences regarding comprehension-related skills or abilities. For theory building, however, it may also be advantageous to establish causal relationships between constructs, and an experiment in the narrower sense of Shadish, Cook, and Campbell (2010) is known to be an efficient way to establish such relationships. Fifth, the tripartite model is flexible with regard to text format; it is not restricted to written text and can thus be easily integrated into theories of multimedia comprehension (Mayer, 2009; Schnotz, 2014). For that reason, it may also constitute a useful framework with reference to the second research goal of investigating the impact of illustrations on children's comprehension of auditory and written text.

2.1 Levels of representation

The tripartite model of text comprehension (van Dijk & Kintsch, 1983) is a constructivist approach holding that three levels of representation can be distinguished as results of text processing: a *text surface* that refers to the memory of the wording, a *textbase* that constitutes the representation of the meaning of larger text units (e.g., sentences or paragraphs) and can be expressed as propositions, and a *situation model*, which is a coherent representation of the state of affairs to which the text refers. Importantly, the situation model is not a mere extraction of explicit information from the text, but a *construction* combining textual information with the recipient's previous knowledge and experience (see also Kintsch, 1988, 1998). This process of construction corresponds to actual text comprehension.

Arguably, typical measurements used in the reading comprehension literature could be appropriated with these three levels of representation. First, memory for text assessed via literal questions could be likened to the text surface. Questions that address meaning at the level of longer text units (sentences or paragraphs), by contrast, require a deeper representation of the kind found in the textbase; this becomes clear when considering identical words with different meanings (e.g., *fair* as adjective and substantive) or – in the case of auditory text – different words with identical pronunciations such as *fair* and *fare*. According to the lexical quality hypothesis (Perfetti, Yang, & Schmalhofer, 2008), a word's meaning has to be derived from its linguistic context in both of the examples above. With regard to *inference* questions, it depends on whether the relevant information can be taken directly from the text or requires background knowledge. For example, when reading or hearing "Debbie changed and wrapped her swimming costume in her towel. She put the bundle in her rucksack" (Cain & Oakhill, 1999, p. 495), the reader must, for information integration, infer across sentences that "she" refers to Debbie and "bundle" refers to the swimming costume in the towel; this can be achieved exclusively with information provided by the text. By constrast, when exposed to a sentence like "Tom and his two sisters fetch their sleeping bags and ground pads from the basement" and asked what these three are going to do tonight, the respondent might be helped by knowing that sleeping bags and ground pads are typical camping equipment so that "camping" is likely the correct answer. Therefore, the former example, known as text-connecting inference (Cain & Oakhill, 1999), corresponds to the textbase level because it requires the integration of information within the text, whereas the latter, which Cain and Oakhill call gap-filling inference, can be subsumed under the situation model because it requires integrating information provided by the text with previous knowledge.

In studies that use these types of questions, Oakhill and colleagues have elaborated clear distinctions between memory for text measured through literal questions, text-based integration tested by text-connecting inferences, and incorporating information outside the text assessed by the use of gap-filling inferences. Cain and Oakhill (1999) found that reading comprehension ability among 7 year-old children is significantly predicted by inferences, especially text-connecting ones, but not by memory for text. Oakhill et al. (2003) have established inference making (both text-connecting and gap-filling inferences), awareness of text structure (e.g., what the title says about a story), and comprehension monitoring as substantial contributors to individual differences in reading comprehension among 8 year-old children. Similarly, results from a longitudinal study (Oakhill & Cain, 2012) suggest that inference-making and comprehension-monitoring skills at age 8 predict reading comprehension two years later even after controlling for the stability of the reading comprehension skill between the two measurement times. Perfetti et al. (2005) have proposed knowledge and retrieval of word meaning (i.e., vocabulary) and word-to-text integration[1] as low-level predictors of comprehension; they suggest that inference making, text structure awareness, and comprehension monitoring are high-level predictors of comprehension (see also Verhoeven & Perfetti, 2008).

Thus far, it could be argued that the reading comprehension components outlined in the previous paragraph have equivalents in the tripartite model; the case of text structure awareness,

[1] Another predictor is word identification (decoding), which contributes to reading, but not listening comprehension and is therefore not addressed in the present study.

however, is less clear. Even so, there are two good reasons to subsume this element under the situation model. One is that knowledge about text structures such the function of the title, the beginning, or the end of a story is required; this knowledge is usually not provided by the text itself. The other is that structural awareness should help text recipients gain an idea what the text is about (regarding the function of a title, this has well been demonstrated by Bransford and Johnson, 1972), which comes closer to the situation model (i.e., a coherent representation of the state of affairs) than to other representation levels. Knowledge and awareness of text structure were beyond the scope of the present research; however, they were controlled for in the experiments by using a text genre with which children should be familiar at school entry: narrative text featuring everyday situations of children in that age group. Table 1 illustrates the processes identified as relevant by influential researchers in the reading comprehension domain (Oakhill, 2020; Perfetti et al., 2005; Verhoeven & Perfetti, 2008) and their classification into the tripartite model (van Dijk & Kintsch, 1983).

Table 1
Processes of text comprehension according to notable research on the development of reading comprehension, and their classification into the tripartite model of text comprehension

Tripartite Model (van Dijk & Kintsch, 1983)	Processes relevant to comprehension as proposed by influential accounts of the development of reading comprehension*
(not classified)	Word identification (decoding)
Text surface	Access to word meanings (vocabulary) Literal memory
Textbase	Word-to-text integration Sentence-based meaning Integration across sentences Text-connecting inferences
Situation model	Gap-filling inferences Integration with prior knowledge Awareness of text structure
(not classified)	Comprehension monitoring

Note. *Cain and Oakhill (1999), Oakhill et al. (2003), Oakhill and Cain (2012), Perfetti et al. (2005), and Verhoeven and Perfetti (2008).

This classification has at least two limitations. First, it is of course, far from covering *all* scientific research conducted on text comprehension development; such an effort would vastly exceed the scope of the present research. Second, the tripartite model does not encompass comprehension monitoring as a meta-cognitive process related to text comprehension. As this process (or related skill) goes beyond the representations of a single text, it appears ill-suited to being subsumed under the levels of text representation. Nevertheless, the tripartite model is a reasonable approach to text comprehension because it can be demonstrated that each level of representation reflects the different processes and related skills that are involved in comprehension. It is also a parsimonious account, at least compared to one involving a host of comprehension processes that have to be distinguished from one another. Although the tripartite model has been addressed explicitly in the reading comprehension literature (e.g., Perfetti et al., 2005), there is, to my knowledge, no empirical work in this research tradition that has systematically applied that model to the development of comprehension. Therefore, doing so is an innovative feature of the present work. The next section describes a method that makes the simultaneous assessment of the tripartite representations possible, thus rendering the tripartite model useful for an experimental investigation of text comprehension like that undertaken here.

2.2 Sentence recognition method

For experimental work on text comprehension, a sentence recognition paradigm has been developed to enable the simultaneous assessment of text surface, textbase, and situation model representations (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986). Participants read or hear a text and are presented with single sentences after a short delay. They are asked to indicate whether these sentences are part of the text they have just read or heard. Each sentence is presented either as an original sentence, requiring a positive answer, or as a *paraphrase*, where the exact wording – but not the meaning – of the sentence has changed (e.g., by using synonyms, scrambling the word order, or switching between active and passive voices), as a *meaning change*, where the semantic structure of the sentence is different while being true to the state of affairs referred to in the text, or as a *situation change*, which is also incompatible with the state of affairs (see Table 2 for a sample sentence).

Schmalhofer and Glavanov (1986) used this method to scrutinize the comprehension of a programmer's manual in a sample of adults. Their results suggest that text recipients discriminate between original sentences and paraphrases; they exhibit a significantly lower proportion of positive answers to paraphrases than to original sentences, indicating that they form a representation of the text surface. Likewise, participants distinguished both paraphrases from meaning changes, indicating that they form a textbase representation, and meaning changes from situation changes, indicating that they also construct a situation model. These findings have been replicated by Fletcher and Chrysler (1990) with narrative text with an adult sample. Nieding (2006) has applied this method to a sample of children aged 5, 8, and 11, who listened to a set of stories with 6 sentences each and then performed the sentence recognition task. A result pattern was obtained that not only suggests that all three representation levels are present in children aged 5 and above, but also that children build these representations regardless of their level of domain-specific knowledge and whether a title that prompts coherence building is provided.

The findings outlined above have two implications: on one hand, they validate the tripartite model (van Dijk & Kintsch, 1983), suggesting that to a certain extent, text recipients memorize the wording (i.e., form a text surface representation), form a semantic representation of what the text explicitly states (textbase), and construct a coherent representation of the state of affairs addressed in the text (situation model). On the other, there is evidence that the sentence recognition method proposed by Schmalhofer and Glavanov (1986) is a valid measure of these three levels of text representation. In recent experiments conducted by our research group, this paradigm was used to compare the representations of different text formats among children (Wannagat, Waizenegger, Hauf, & Nieding, 2018; Wannagat, Waizenegger, & Nieding, 2017). In Wannagat et al.'s (2017) study, for example, representations of auditory, audiovisual, and written text were compared among children aged 8 and 10 years as well as adults. The main results indicated that situation model construction based on written text improves substantially with age; the 8 year-old children demonstrated more accurate situation models with auditory and audiovisual compared to written text, whereas the older participants exhibited comparable results across text conditions. These results, along with those obtained by Wannagat et al. (2018) and outlined in the next section, indicate that text comprehension among children depends on the text presentation format. In summary, the construct validity of the tripartite model and the sentence recognition method and their evident usefulness for experimental research are the main reasons why they were also used in the experiments discussed in this thesis.

2.3 Comprehension of text and pictures

As this thesis examines the influence of illustrations on children's comprehension of auditory and written text, this section begins by outlining early empirical and theoretical work in the field of multimedia comprehension before turning to a framework that uses the tripartite model to explain these insights. This framework, the ITPC (Schnotz, 2014; Schnotz & Bannert, 2003), provides an underpinning for the research reported here.

Several empirical attempts have been made to compare children's comprehension of auditory stories with and without dynamic illustrations (i.e., radio vs. television; Beagles-Roos & Gat, 1983; Gibbons, Anderson, Smith, Field, & Fischer, 1986; Gunter, Furnham, & Griffiths, 2000; Hayes, Kelly, & Mandel, 1986; Pezdek et al., 1984; Walma van der Molen & van der Voort, 2000), auditory stories with and without static illustrations (e.g., Greenhoot & Semb, 2008; Guttmann, Levin, & Pressley, 1977), and written stories with and without static illustrations (Gambrell & Jawitz, 1993; O'Keefe & Solman, 1987; Pike, Barnes, & Barron, 2010). Typical comprehension measures have been free recall, cued recall, recognition of central and peripheral story elements, and inference making. A large portion of these results indicates higher performance under illustrated than under non-illustrated text conditions (Beagles-Roos & Gat, 1983; Gambrell & Jawitz, 1993; Gibbons et al., 1986; Greenhoot & Semb, 2008; Pezdek et al., 1984; Pike et al., 2010). However, children appear to recall more expressive language and draw more inferences when provided with verbal text alone (Beagles-Roos & Gat, 1983; Hayes et al., 1986). A few of these investigations (Gibbons et al., 1986; Pike et al., 2010) have developmental implications; namely that younger children (4–5 year-old children in Gibbons et al.'s study, 7–9 year-old children in Pike et al.'s work) benefit more from illustrations than older ones (7 year-olds for the Gibbons group, 11 year-olds for the Pike group; however, Greenhoot and Semb reported the opposite pattern of results in a sample of 4 and 5 year-old preschoolers). Taken together, these studies suggest the superiority of illustrated text for most but not all indicators of narrative text comprehension among children.

In the domain of learning from expository text, Mayer (1997, 2009) has proposed the *multimedia principle*, which holds that text recipients learn better from verbal text accompanied by pictures than from verbal text alone. This notion is based not only on a large body of empirical findings (for a review, see Carney & Levin, 2002) but also on well-established theories of cognition and memory; one is the multicomponent working memory model assuming independent processing subsystems for visual-spatial and verbal-articulatory content (e.g., Baddeley, 2002). As a consequence, the limited capacity of working memory can be used more efficiently for comprehension-related processing when both verbal and visual-spatial information is available. Paivio's (1986) dual coding theory, according to which verbal and nonverbal information is encoded and manipulated in different but interconnected memory sections, is a similar approach. Accordingly, pictures can provide additional memory traces that may support the comprehension of verbal text. The multimedia principle appears to apply to a variety of learning forms and text formats, as Butcher (2014) points out in her review. According to several studies (e.g., Butcher, 2006; Cuevas, Fiore, & Oser, 2002; Glenberg & Langston, 1992), pictures improve performance on indicators of superficial and especially deeper forms of learning.

2.3.1 Integrated model of text and picture comprehension

The ITPC (Schnotz, 2014) describes the comprehension of text-picture units in more detail and roughly distinguishes between two stages of processing. During the early *perceptual surface structure processing* stage, the modality-specific sensory registers (eyes and ears) transmit verbal and pictorial text input to the working memory through the visual and auditory channels. The second stage, *semantic deep structure processing*, serves as an explanation for the multimedia principle (Mayer, 2009) with the aid of the tripartite model (van Dijk & Kintsch, 1983). This stage involves two different processing paths, one descriptive and one depictive. At the beginning of descriptive processing, a text surface representation is formed that can include spoken or written text, with the latter being transferred into phonological lexical patterns. In the next step, a textbase is formed via semantic processing (parsing). Finally, a mental (situation) model is constructed by integrating the textbase with lexical, perceptual, and conceptual knowledge retrieved from long-term memory (cf. Ericsson & Kintsch, 1995).

However, according to the ITPC, the situation model itself is a depictive representation of text and can also be obtained directly on the depictive processing path via *analog structure mapping* (Gentner, 1989). This process is based on the picture surface representation that is gathered on the depictive path (analogously to the text surface) and is due to structural similarities between picture surface representations and situation models. Schnotz and Bannert (2003) have demonstrated empirically that analog structure mapping can be used to match a previously constructed situation model with the picture surface representation and thus strengthen the situation model's memory trace. Finally, the ITPC proposes that *model inspection* can take place when an existing situation model is updated in line with newly integrated verbal or pictorial information. In the course of model inspection, the updated situation model is encoded in propositional (i.e., textbase) format. According to Schnotz and Bannert (2003), the main function of model inspection consists of preparing verbal utterances. As a consequence, this implies that the original textbase representation can be altered by nonverbal information, including illustrations, that is encoded after the construction process.

Recently, Wannagat et al. (2018) compared auditory with audiovisual stories in a sample of 7 to 11 year-old children using Schmalhofer and Glavanov's (1986) sentence recognition method. They reported significantly higher situation model and text surface representations when illustrations are present rather than absent, while a reverse effect was observed for the textbase. Wannagat et al. have argued that analog structure mapping may make a considerable contribution to a stronger representation of the situation model and that, at the same time, fewer cognitive resources are used for the construction process itself. They propose that this results in a weaker representation of the textbase that marks the interim step between the text surface and the situation model on the descriptive path. It has been further suggested that these freed-up resources may be used to memorize the wording of the text, which then results in higher text surface representations in the audiovisual than in the solely auditory text condition. In sum, Wannagat et al. (2018) have demonstrated that the ITPC (Schnotz, 2014), which was originally designed for learning from expository text, can also explain results related to the comprehension of illustrated narrative text.

2.3.2 Influence of illustrations on the comprehension of written text

As the research in this thesis deals with the effect of illustrations on the comprehension of both auditory and written narrative text, another specification of Mayer's multimedia principle, referred to as the *modality principle* (Low & Sweller, 2014; Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995), may be of particular relevance. It suggests that multimedia learning benefits more from a text that employs two different sensory channels (auditory-visual) instead of one (visual-visual), which leads to the expectation that illustrations added to an auditory story have a more beneficial effect on comprehension than those added to a written one. The modality principle can be explained within the ITPC framework (Schnotz, 2014), specifically in the context of perceptual surface structure processing: as the capacity of each sensory channel is limited, more input can be processed when the text includes both auditory and visual elements rather than solely visual ones.

Moreover, written text with pictures necessarily requires early (but not late) steps of text processing to take place successively. From a practical perspective, this raises the question of whether it is better for comprehension to start with the written text or with the picture. In the domain of expository text, Eitel and Scheiter (2015) conducted a systematic review of studies examining this issue. The results they report are heterogeneous, to say the least; some indicate better comprehension when the verbal text is presented after the picture (e.g., Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013) whereas others indicate the opposite pattern (e.g., Canham & Hegarty, 2010). Eitel and Scheiter (2015) refer to a number of theoretical accounts, one of which is the ITPC, to integrate these divergent findings. Accordingly, pictures may serve as a scaffold for understanding subsequent verbal text and thus enhance situation model construction, whereas pictures that follow verbal text may interfere with a mental model that has already been constructed from verbal text via model inspection (Schnotz, 2014). As far as I know, O'Keefe and Solman's (1987) set of experiments is the only case in the area of narrative text comprehension where the order of written text and pictures has been systematically varied. A combined analysis of their first two experiments reveals that story comprehension improves when illustrations are presented before or after their corresponding verbal text compared with verbal text only, but there appears to be no difference that can be attributed to the order of verbal text and illustrations. Therefore, the first experiment in the present study has been designed to shed more light on this issue.

2.3.3 Static versus dynamic illustrations

Although several studies have compared the comprehension of auditory and audiovisual stories, evidence from direct comparisons between audiovisual stories with static as opposed to dynamic illustrations remains scarce. In the context of expository text, however, there are a considerable number of comparable investigations: Höffler and Leutner (2007) performed a meta-analysis based on 76 pair-wise comparisons of animated versus static illustrations for scientific text learning. The overall effect size leans moderately ($d = 0.40$) toward animated illustrations. It is noteworthy, however, that the results reported in this meta-analysis are quite heterogeneous; some studies have found that learning is better when dynamic rather than static illustrations are used (e.g., Catrambone & Seay, 2002; Rieber, 1990), whereas others reveal the opposite pattern (e.g., Mayer, Hegarty, Mayer, & Campbell, 2005). More importantly, theoretical underpinning is lacking: as Höffler and Leutner (2007) have themselves pointed out, neither the cognitive theory of multimedia learning (Mayer, 2009) nor the ITPC (Schnotz, 2014) explicitly distinguishes between static and animated pictures. Therefore, it appears useful to examine the cognitive processes involved in the comprehension of text-illustration units more thoroughly so that differences between static and animated illustrations in that context may also be explained. Following this reasoning, an embodied account of the situation model is introduced in the next section.

2.4 Grounding the situation model in perception and action

Early accounts of the situation model (e.g., Kintsch, 1988, 1998; van Dijk & Kintsch, 1983) conceptualize it as a connection between propositions based on verbal text input (textbase) and propositions based on information retrieved from long-term memory. For this constructivist research tradition, the question of whether situation models are based on perceptual input appears to be of minor relevance. However, since the 1990s, another line of research has emerged which holds that situation models are externally grounded in perception and action (Glenberg & Robertson, 2000; Zwaan, 1999). This proposal originates from the symbol grounding problem (e.g., Harnad, 1990), which reflects the notion that symbolic representations of the world such as words cannot carry meanings unless they ultimately have a non-symbolic (sensory) relationship with the world. In this context, Barsalou's (1999) *embodied cognition* theory is particularly influential; fundamentally, it assumes that memory, language, and thought are rooted in internal representations of sensorimotor or introspective neural experience. These representations are modal, which means that they are processed through the same neural systems as the perceptual states that produce them. They can also be integrated into an organized system so that a situation can be re-experienced in its absence. This process of *perceptual simulation* is assumed to take place routinely when dealing with symbolic representations of the world, including text comprehension. Analogously, a process called *motor resonance* has been proposed to describe the simulation of actions through the motor system during text reception (Glenberg & Kaschak, 2002; Taylor & Zwaan, 2008; Zwaan & Taylor, 2006; Zwaan, Taylor, & de Boer, 2010). In line with these assumptions, the situation model can be regarded as an analogous, multimodal simulation of the situation described in the text that resembles the mental representation of real-life experience (Zwaan, 1999, 2014).

2.4.1 Perceptual simulation of static and dynamic object features

Since Barsalou's (1999) foundational work, efforts have been made to test the perceptual simulation hypothesis empirically, with many studies involving objects that occur in a sentence and thus should be part of the situation model. Stanfield and Zwaan (2001) designed a *picture verification task* for this purpose. Their participants read a sentence that included a target object whose implied spatial orientation was either horizontal (e.g., "Liz hammered the nail into the wall") or vertical (e.g., "Liz hammered the nail into the ceiling"). After reading, participants saw the picture of an object and had to decide as quickly as possible whether the object had been mentioned in the sentence. In trials where the picture showed the target object ("nail"), the orientation of the depicted object either matched or mismatched the orientation implied by the sentence. The rationale behind this procedure was that if participants perceptually simulated the object's orientation implied by the sentence, they would exhibit shorter response times in matching than in mismatching trials. Indeed, the adult participants in Stanfield and Zwaan's (2001) study demonstrated precisely this result pattern, which indicates that they perceptually simulated an object's orientation during text comprehension. Further studies with similar procedures have gathered considerable evidence for the perceptual simulation of different static object features. Besides orientation, these include shape (Zwaan, Stanfield, & Yaxley, 2002), size (de Koning, Wassenburg, Bos, & van der Schoot, 2017), color (Connell, 2007; Richter & Zwaan, 2009; Therriault, Yaxley, & Zwaan, 2009), distance (Winter & Bergen, 2012), number (Patson, George, & Warren, 2014), and visibility (Horton & Rapp, 2003; Yaxley & Zwaan, 2007).

Regarding the perceptual simulation of dynamic object features, a number of studies (e.g., Bergen, Lindsay, Matlock, & Narayanan, 2007; Kaschak et al., 2005; Kaschak, Zwaan, Aveyard, & Yaxley, 2006; Richardson, Spivey, Barsalou, & McRae, 2003; Zwaan, Madden, Yaxley, & Aveyard, 2004) have provided evidence that adults also simulate an object's direction of movement. Zwaan et al.'s (2004) participants listened to sentences that implied an object moving either towards ("Bob hurls the softball at you") or away ("You hurl the softball at Bob") from their body. Then, they saw a rapid sequence of two pictures, the second of which was either larger or smaller than the first, giving them the impression of movement toward or away from them. The participants were instructed to decide whether the two pictures showed the same object. The results indicated shorter response times when the direction described in the sentence matched the direction implied by the picture sequence and longer response times when they did not match. Other movement characteristics, such as speed, have also been found to be perceptually simulated by adults (e.g. Speed & Vigliocco, 2014).

2.4.2 Developmental research on perceptual simulation

Compared with research involving adult participants, there are only a few studies of perceptual simulation during childhood development. Engelen, Bouwmeester, de Bruin, and Zwaan (2011) found that the spatial orientation and shape of objects are perceptually simulated by 8 to 12 year-old children; similar results were obtained by de Koning et al. (2017) regarding size. Finally, Hauf (2016) has demonstrated that 6 year-old children perceptually simulate the shape of an object implied by a story to which they have listened. These results suggest that the perceptual simulation of static object features is largely developed by age 6. Regarding the perceptual simulation of movement, there are results suggesting that the simulation of speed might have already developed in early childhood. In Fecica and O'Neill's (2010) study, 3 to 5 year-old children listened to a story featuring a child protagonist who either walked (slow condition) or was being driven (fast condition) somewhere. The story was presented sentence by sentence, and the participants determined the speed of listening by proceeding with mouse clicks from one sentence to the next. Participants in the fast condition revealed a higher listening speed than participants in the slow condition, indicating that children simulate movement speed. The perceptual simulation of movement direction among children has been observed in a picture verification study (Hauf, Nieding, & Seger, 2020): children between 6 and 10 years of age and adults listened to sentences that included an object moving up or down (e.g., "The apple falls on the grass"). Then, an animated picture of the target object appeared moving to the center of the screen, either from the bottom or the top. Both children and adults appeared to be faster in matching (e.g., apple dropped from the top) than in mismatching (e.g., apple rose from the bottom) trials, suggesting that the perceptual simulation of movement is at least broadly developed among children by their last year of preschool.

2.4.3 Motor resonance

The motor resonance hypothesis has been tested empirically by associating either the text reception itself (e.g., reading) or a text-related task (e.g., sensibility judgment) with a manual response that either matches or mismatches a movement described in the text. For example, Zwaan and Taylor (2006) worked with sentences featuring a clockwise rotation (e.g., "Frank turned up the volume") or a counterclockwise one (e.g., "Frank turned down the volume"). In the second experiment of that study, adult participants judged whether a sentence of this type made sense by turning a knob either clockwise or counterclockwise. The obtained response times were significantly faster when the rotation implied by the sentence matched the rotation associated with the correct response than when they did not match. The participants in that study's fourth experiment turned a knob in order to read sentences of the same kind word by word. As expected, reading times were significantly faster when the rotation response during reading matched than when they did not match the rotation implied by the sentence. Similar results were previously obtained by Glenberg and Kaschak (2002) using movements toward versus away from the observer. In addition, neuroimaging studies have yielded an association between action-related language processing and activation in the same motor and premotor areas of the brain as those associated with overt action (e.g., Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005) and action observation (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006). This appears to pertain especially to actions that involve the hands, feet, and mouth.

In sum, there is ample solid evidence that text reception involves embodied cognition processes such as perceptual simulation and motor resonance among adults; there are also some results suggesting that children in their early school years perceptually simulate the static and dynamic features of the situation referred to in a text. For this thesis, however, it is of central interest whether these embodied processes are relevant for text comprehension and, more specifically, for the comprehension of narrative text that does (or does not) include illustrations. Some theoretical and empirical approaches to this issue are addressed below.

2.4.4 Significance of perceptual simulation and motor resonance for text comprehension

Early theoretical work in the field of embodied cognition holds that perceptual simulation underlies situation models (Barsalou, 1999) and that perceptual simulation provides a vast amount of information that amodal symbols simply cannot achieve, but that might be relevant for the construction of an appropriate situation model (Zwaan, 1999). This assumption has been supported in two experiments reported by Kaschak et al. (2005), whose participants listened to sentences that included vertical movements (up or down) or movements toward or away from them. During listening, the participants were exposed to animated visual stimuli that were unrelated to the verbal content and showed either a matching or a mismatching movement. The results indicate that matching stimuli *impair* performance in subsequent sensibility (experiment 1) and grammaticality (experiment 2) judgment tasks. The authors suggest that matching visual stimuli presented parallel to the text engage neural mechanisms that may otherwise be involved in perceptual simulation. Given that the suppression of perceptual simulation is associated with lower performance in comprehension-related tasks, it is reasonable to conclude that it plays a facilitating role in comprehension. Interestingly, this insight appears to pertain not only to task types that involve situation model construction (i.e., sensibility judgment) but also to a grammaticality judgment task that one might expect to be solved at the textbase level.

However, this line of evidence has since proven tenuous: the third experiment in Zwaan and Taylor's (2006) study involved sensibility judgments of sentences that described rotation movements. During listening, participants had to watch a fixation cross that rotated in a direction that either matched or mismatched the rotation direction implied by the sentence. Consistent with the results obtained in Zwaan and Taylor's second and fourth experiments, a matching advantage has emerged that, according to the authors, provides support for the motor resonance hypothesis. Nevertheless, this can also be seen as a failure to replicate the results obtained in a similar experiment; namely, the first experiment in Kaschak et al.'s (2005) study, where a *mis*-matching advantage was observed. Other than experiments with subsequent matching versus mismatching stimuli, where matching stimuli have been associated with more efficient task performance, experiments with parallel (i.e., possibly interfering) stimuli have yielded more heterogeneous results. As a consequence, these results are far from unanimously indicating that perceptual simulation is functional to text comprehension; it thus appears that additional and more heterogeneous empirical work is needed to obtain more insights into this issue.

As noted above, the more specific goal of the present work is gauging the functional role of embodied cognition in the comprehension of *illustrated* narrative text. In one of our earlier studies that established more accurate situation models in illustrated than in non-illustrated stories (Wannagat et al., 2018), we argued that an illustration added to a story resembles a real-life experience (Barsalou, 2008) and may thus be able to initiate and support perceptual simulation processes. Under this logic, analog structure mapping between internal visual representations and surface representations of external pictures (Gentner, 1989; Schnotz & Bannert, 2003) may therefore not be restricted to a structural correspondence that can be described in terms of propositions; rather, it may extend to a modal correspondence grounded in perception. In short, this would imply that perceptual simulation boosts the situation model in illustrated text. A necessary but not sufficient premise of this assumption is that perceptual simulation is more pronounced when the text includes illustrations than when it is presented solely in written words. One of the experiments referred to in this thesis addresses this premise. Analogously, one might expect that dynamic pictures also initiate and support motor resonance and that motor resonance may then bolster the situation model, a claim that Taylor and Zwaan (2009) have discussed in detail. Therefore, another experiment reported here was designed to provide more insight into this topic.

2.5 Perceptual boundedness and perceptual support

The main purposes of this research are understanding the development of text comprehension during childhood and determining the impact of adding illustrations to verbal text on children's text comprehension. Of course, these two research questions can also be addressed in a combined manner: how does the *comprehension of illustrated versus non-illustrated text develop during childhood*? This combined question deserves more investigation, given that established theories of illustrated text comprehension like the multimedia principle (Mayer, 2009) and the ITPC (Schnotz, 2014) are fairly insensitive to development, whereas prevalent accounts of text comprehension among children (mostly reading comprehension; see, e.g., Perfetti et al., 2005) often do not take different presentation formats of text into account. Therefore, a theoretical framework that asks how children use perceptual information during text reception in the course of development may be useful.

There is some evidence that younger children rely more on perceptual information during text reception than older children and adults (e.g., Gibbons et al., 1986; Pike et al., 2010). In a study of predictive visual inferences during narrative text comprehension, such as whether one creates the image of a broken window while hearing about someone who kicks a ball against it (Unsöld & Nieding, 2009), only the youngest group of 6 year-old preschoolers exhibited significantly shorter response times to predictive stimuli (e.g., a broken window) than to non-predictive ones (e.g., an intact window) in a picture-naming task. Interestingly, this result has only been observed with audiovisual stories; in solely auditory versions of these stories, no evidence for predictive inferences was obtained for any age group. As an explanation, the authors suggest that preschoolers overrepresent perceptual information during situation model construction, whereas older children and adults are more parsimonious in this regard, meaning that their situation models do not include task-irrelevant inferences. This overrepresentation of perceptual information among younger children is sometimes referred to as *perceptual boundedness* (Springer, 2001).

On the basis of Unsöld and Nieding's (2009) results, it is worth considering that illustrations may help younger children with situation model construction so that cognitive resources can be spared; these resources can then be used for drawing predictive inferences. This is in line with Springer's (2001) *perceptual support* hypothesis, according to which perceptual information can serve as a scaffold for cognitive development during childhood. More specifically, this concerns the development of concepts, where a shift from relying on characteristic features to relying on defining ones has been observed (e.g., Keil & Batterman, 1984). Arguably, this shift follows an adaptive strategy; namely, that characteristic features – which, according to Keil and Batterman (1984), correspond closely with perceptually salient features – can be relied upon as long as there is insufficient background knowledge for the construction of categorical concepts based on defining features. As there is usually a positive correlation between characteristic and defining features of members of the same category, it can also be argued that overreliance on perceptually salient features in early childhood supports the development of definition-based concepts. Springer (2001) outlines different specifications of the perceptual support theory to explain how perceptual information contributes to cognitive development. The realist argument proposes that perceptual information *per se* is sufficient; however, as Springer notes, this argument has proven empirically untestable.

A more promising account of Springer's (2001) perceptual support hypothesis is the fluency argument, which suggests that perceptual information serves as an input for the practice of certain cognitive skills before these skills are sophisticated enough to be applied to non-perceptual information. Notably, these skills include analog structure mapping, which has been suggested earlier in the present work to be relevant for the comprehension of illustrated text (cf. Gentner, 1989; Schnotz, 2014). Ruggeri and Katsikopoulos (2013) tested the perceptual support hypothesis in a sample of children aged 7 to 10 and in adults. They used a single-choice decision task (e.g., which one of two cars is more expensive) in which participants were instructed to use cues that allowed them to draw relevant inferences. In the first experiment, where the participants generated their own cues, the decisions made by the younger group (aged 7–8 years) were at least as accurate as those made by the older children and adult groups by using cues based on perceptually salient features (such as the length and width of the car). By contrast, in the second experiment, where non-perceptual cues (e.g., horsepower) were provided by the experimenter, the younger children's decisions were less accurate than those of the older participants. Both results indicate that perceptual information can help younger children develop an accurate concept of real-world features (such as car prices) as long as they do not have enough background knowledge to rely on non-perceptual information. In the field of narrative text, research on illustrated storybooks read to preschoolers (e.g., Isbell, Sobol, Lindauer, & Lowrance, 2004) provides evidence that storybook reading improves the development of text comprehension skills and that young recipients rely heavily on the illustrations during story retelling. Although this might be mere coincidence, the perceptual boundedness and perceptual support hypotheses can serve as a theoretical starting point for research on the comprehension of text-picture units from a developmental perspective.

According to recent research on the role of embodiment in conceptual development and language learning (e.g., Pexman, 2019), there is evidence that early concepts emerge from sensorimotor experience, whereas concepts that emerge later during development are less strongly tied to that experience. Pexman's (2019) findings regarding the role of embodied processes in conceptual learning are in line with Springer's (2001) perceptual support hypothesis: embodied cognition processes like perceptual simulation or motor resonance may be related to the processing of perceptual information, including pictures and animations, which in turn helps younger children with their conceptual development. Studies on novel verb learning (e.g., Hald, van den Hurk, & Bekkering, 2015) have revealed that children up to 8 years of age are more successful when they have the opportunity to simulate the action associated with a novel verb through their sensory and motor systems than when they do not. Therefore, it would be valuable to know whether the embodied cognition processes that occur during text comprehension are also subject to perceptual support. Hauf's (2016) experiments on the perceptual simulation of object shape during narrative text comprehension provide evidence in this regard. Perceptual simulation occurred among 6 year-old children but not among older children or adults when the stories were presented in an audiovisual format. By contrast, neither children nor adults exhibited perceptual simulation when auditory-only versions of the same stories were provided. These results are consistent with the view that perceptual support may be relevant for the comprehension of illustrated text. First, it can be argued that illustrations provide 6 year-old children with perceptual support and that these younger children process this perceptual information through their sensory systems. Second, older children and adults do not use their cognitive resources for perceptual simulation because they no longer benefit from perceptual support. Finally, when illustrations – and thus, perceptual support – are not provided, even 6 year-old children do not benefit from perceptual simulation; as a consequence, they do not simulate and thus spare their cognitive resources.

3. Rationale and Research Methodology

The first aim of the empirical work reported here was to examine how children in their early school years understand written and auditory text and to identify the developmental changes related to text comprehension that occur within this age range. In accordance with the tripartite model (Kintsch, 1988, 1998; van Dijk & Kintsch, 1983), the first research question asked whether children form a text surface and a textbase representation and construct a situation model when reading or listening to a story. Earlier results obtained by our research group (Nieding, 2006; Wannagat et al., 2017, 2018) suggest that they do.

The second research goal was to determine the impact of illustrations on children's representations of auditory and written narrative text. Accordingly, the second research question addressed the impact of illustrations on children's text surface, textbase, and situation model representations. Expectations regarding the impact of illustrations on the text surface were in line with the findings from earlier studies, notably Wannagat et al. (2018). Regarding both the textbase and the situation model, the assumption was considered that illustrations enable the obtaining a situation model on the depictive path (analog structure mapping). The textbase was regarded as secondarily affected by subtle changes to the situation model based on pictorial information; this process is referred to as model inspection. Both the analog structure mapping and the model inspection hypotheses are derived from the ITPC (Eitel & Scheiter, 2015; Schnotz, 2014; Schnotz & Bannert, 2003). With particular regard to the situation model, embodied approaches to comprehension were also taken into account (e.g., Barsalou, 1999; Glenberg & Robertson, 2000; Taylor & Zwaan, 2009; Zwaan, 1999, 2014; Zwaan & Radvansky, 1998), specifically the assumption that perceptual simulation and motor resonance are functionally related to text comprehension.

If illustrations added to a text actually do support the situation model, the role of embodied cognition processes such as perceptual simulation still deserves further investigation. Wannagat et al. (2018) argue that illustrations added to a story initiate and support perceptual simulation, which in turn helps with improving the situation model. This assumption presumes differences regarding the manifestation of perceptual simulation in illustrated versus non-illustrated text. For that reason, a third research question emerged that examined how illustrations affect perceptual simulation.

The final issue to be investigated was whether the cognitive development of children within the age range of interest (5–13 years) has a noticeable influence on the comprehension of illustrated and non-illustrated text. The fourth research question thus asked whether younger children, when compared to older ones, tend to overrepresent perceptual information (perceptual boundedness) or to use perceptual information to practice comprehension-related skills (perceptual support; see Isbell et al., 2004; Ruggeri & Katsikopoulos, 2013; Springer, 2001).

This research comprised three experiments. The first two assessed text surface, textbase, and situation model representations by using a sentence recognition method (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986) in samples of children aged 7 to 13. In the last experiment, perceptual simulation was assessed using a picture verification task among 5 to 10 year-old children and among adults. After outlining the four research questions and their related hypotheses, this chapter describes these two research methodologies.

3.1 Research Question 1: Three levels of representation

The assumption that children form representations of text surface, textbase, and situation model during text reception is congruent with a necessary methodological premise of the sentence recognition method applied here; namely, that the constructed change types (paraphrases, meaning changes, and situation changes) reflect incompatibility with their respective levels of representation. Accordingly, acceptance rates – the proportion of positive answers to the question of whether a sentence was part of the text – should be highest for original sentences and lowest for situation changes, with a significant graduation between all change types. Research question 1 is thus divided into the following hypotheses, which are investigated in Experiments 1 and 2:

> *Hypothesis 1a.* The acceptance rate for original sentences is significantly higher than that for paraphrases.
>
> *Hypothesis 1b.* The acceptance rate for paraphrases is significantly higher than that for meaning changes.
>
> *Hypothesis 1c.* The acceptance rate for meaning changes is significantly higher than that for situation changes.

3.2 Research Question 2: Impact of illustrations on text comprehension

Here, it is specified how illustrations are expected to affect children's text surface, textbase, and situation model representations of narrative text. For the situation model, there is empirical evidence from previous research, notably Wannagat et al. (2018) that illustrations support situation model construction from auditory verbal text. Earlier studies with narrative text also suggest that deeper forms of comprehension benefit particularly from illustrations added to an auditory (e.g., Gibbons et al., 1986; Hayes et al., 1986; Ricci & Beal, 2002) or written story (e.g., Gambrell & Jawitz, 1993). The ITPC holds that when pictures are present, situation models can be directly obtained through analog structure mapping using the depictive path (Gentner, 1989; Schnotz, 2014). If this depictive path converges with the construction process on the descriptive path, as should be the case when the text is illustrated properly (cf. Schnotz & Bannert, 2003), the situation model representation may be stronger than in text formats, where only one path is available. Therefore, the following hypothesis refers to Experiments 1 and 2:

> *Hypothesis 2a.* Situation model representations are more precise when verbal stories are presented with illustrations rather than without.

According to Eitel and Scheiter (2015), analog structure mapping based on a picture may serve as a scaffold for subsequent situation model construction based on verbal text, with the latter presumed to be a more complex process than the former. This was examined in Experiment 1, where the order of written narrative text and illustrations was systematically varied:

> *Hypothesis 2b.* Situation model representations are more precise when illustrations are presented before rather than after their corresponding written stories.

Another specific hypothesis with respect to the situation model originated from Wannagat et al.'s (2018) suggestion that analog structure mapping may be induced by perceptual simulation processes. If this is true for both static and dynamic illustrations, one could analogize that the simulation of actions through the motor system (motor resonance; see Taylor & Zwaan, 2009) would be induced by dynamic illustrations in particular. Therefore, one hypothesis claims that animated illustrations further enhance the situation model by reinforcing motor resonance and thus supporting the analog structure mapping of dynamic information; this supposition was examined in Experiment 2:

> *Hypothesis 2c.* Situation model representations are more precise when dynamic rather than static illustrations are presented with auditory stories.

Semantic information (textbase) has been found to be affected negatively by illustrations in a previous study comparing auditory and audiovisual stories (Wannagat et al., 2018). The ITPC can account for this result. The availability of analog structure mapping in the presence of pictures diminishes the relevance of situation model construction on the descriptive path so that the textbase, which is the direct precursor of the situation model on this path, is represented less strongly. Experiments 1 and 2 were intended to replicate this result:

> *Hypothesis 2d.* Textbase representations are more precise when stories are presented without rather than with illustrations.

The ITPC further claims that model inspection, which refers to the updating of semantic information based on changes to the situation model, may substantially alter the textbase representation (Schnotz & Bannert, 2003). If the verbal text is processed first, a situation model constructed on the descriptive path may be updated on the basis of a subsequent picture; this update can affect the textbase, resulting in a less accurate representation of the original textbase. Therefore, Experiment 1 addressed the following hypothesis:

> *Hypothesis 2e.* Textbase representations are more precise when illustrations are presented before rather than after their corresponding written stories.

For accuracy on the text surface level (i.e., remembering the exact wording), illustrations were expected to have a positive effect, in line with earlier results indicating that participants recalled or recognized more literal details when provided with illustrations (e.g., Pezdek et al., 1984; Ricci & Beal, 2002; Wannagat et al., 2018). One possible explanation is that fewer cognitive resources are needed for the situation model, leaving more capacity for memorizing literal information. Therefore, Experiments 1 and 2 addressed the following hypothesis:

> *Hypothesis 2f.* Text surface representations are more precise when stories are presented with rather than without illustrations.

To my knowledge, there is no sufficient theoretical or empirical groundwork that would support hypotheses as to how the order of verbal text and illustrations (Experiment 1) or the illustrations' static or dynamic character (Experiment 2) affect the text surface, so these questions were addressed in an exploratory way.

3.3 Research Question 3. Perceptual simulation of vertical movement

According to embodied cognition approaches (e.g., Barsalou, 1999; Zwaan, 1999), text recipients re-experience features of the situation described in the text through their perceptual systems. This phenomenon, known as perceptual simulation, has been demonstrated empirically with static features of the situation in both child and adult samples (e.g., de Koning et al., 2017; Engelen et al., 2011; Hauf, 2016; Patson et al., 2014; Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012). For dynamic aspects of perceptual simulation, however, evidence from research with children is tenuous at best. Recently, Hauf et al. (2020) reported that children between 6 and 10 and adults perceptually simulate vertical object movements during sentence comprehension. Experiment 3 of the present work sought to replicate this result with stories rather than sentences. Therefore, the first hypothesis regarding perceptual simulation is as follows:

> *Hypothesis 3a.* Children and adults perceptually simulate vertical object movements when listening to narrative text.

Even more importantly, Experiment 3 was intended to examine whether perceptual simulation is more pronounced in illustrated than in non-illustrated text, which is a necessary premise of the assumption that perceptual simulation improves the situation model in illustrated as opposed to non-illustrated text (Wannagat et al., 2018). Another novelty of Experiment 3 was therefore that perceptual simulation of movement was compared between auditory and audiovisual stories. Similar results using static object features are reported in Hauf's (2016) doctoral thesis. Children aged 6 years were found to simulate an object's shape when listening to an illustrated story, but not when listening to a non-illustrated one. Therefore, the second hypothesis related to perceptual simulation is as follows:

> *Hypothesis 3b.* Children and adults perceptually simulate vertical object movements more in a more pronounced fashion when listening to audiovisual rather than solely auditory narrative text.

3.4 Research Question 4. Perceptual boundedness and perceptual support

The last research question refers to age-related differences that may have developmental implications. Corresponding hypotheses can be derived from the perceptual support account, which suggests that perceptual cues associated with verbal information are functional to the development of comprehension-related cognitive skills (Ruggeri & Katsikopoulos, 2013; Springer, 2001). On the other side of that coin, younger children – especially preschoolers – appear to rely heavily on pictures during comprehension-related activities like text retelling and inference making (Isbell et al., 2004; Unsöld & Nieding, 2009), a phenomenon referred to as perceptual boundedness (Springer, 2001). Both the perceptual boundedness and perceptual support accounts predict that younger children benefit more from pictures than older ones, especially regarding deeper comprehension such as the situation model. Therefore, the following hypothesis is proposed for Experiments 1 and 2:

Hypothesis 4a. For the situation model, the advantage of illustrated versus non-illustrated stories is negatively related to age.

In line with the perceptual boundedness account, one may also argue that the perceptual simulation of movement is more strongly manifested among younger children than among older ones. This is consistent with Hauf's (2016) examinations of the perceptual simulation of object shapes and with Unsöld and Nieding's (2009) findings concerning predictive visual inferences. In both cases, 6 year-old children, but not older children or adults, were found to overrepresent perceptual information. Therefore, Experiment 3 tested the hypothesis that 6 year-old children manifest perceptual simulation more strongly than older participants:

Hypothesis 4b. When listening to narrative text, 6 year-old children perceptually simulate vertical object movements in a more pronounced fashion than older children or adults.

Following recent findings indicating that language learning among younger but not necessarily older children draws heavily on sensorimotor experiences (e.g., Pexman, 2019), it is reasonable to suggest that embodied cognition processes such as perceptual simulation also play a role in perceptual support in the field of text comprehension. Therefore, it was predicted that perceptual simulation underlies the processing of illustrations and that illustrations thus provide younger children with perceptual support. Accordingly, and in line with earlier findings regarding the perceptual simulation of static object features during narrative text comprehension (Hauf, 2016), 6 year-old children – but not older children or adults – were expected to exhibit perceptual simulation more strongly in audiovisual than in auditory text. This, again, was addressed in Experiment 3:

Hypothesis 4c. When listening to audiovisual rather than solely auditory narrative text, 6 year-old children perceptually simulate vertical object movements in a more pronounced fashion than older children or adults.

3.5 Methodology of Experiments 1 and 2. Sentence recognition

Experiments 1 and 2 of the present work used a sentence recognition task based on Schmalhofer and Glavanov's (1986) paradigm and adapted from Wannagat et al. (2017, 2018). It consisted of 12 stories with six sentences each; these were related to plausible events in the daily lives of children in Western societies (see Figure 1 for a sample story in English and Appendix D1 for the entire set of stories in the original German). During the presentation phase, participants read or listened to these stories sentence by sentence. In illustrated text conditions, one illustration was associated with each sentence; Figure 1 depicts the static illustrations of a sample story. Dynamic illustrations were animated versions of the static ones; these animations were reduced to a minimum and emphasized movements of the human body, especially the leg, hand, and mouth movements that were associated with motor resonance in a neuroimaging study (Aziz-Zadeh et al., 2006). Sample descriptions of movements are also included in Figure 1.

After each block of four stories, a task phase was implemented; participants were exposed to single sentences in scrambled order and had to decide whether each has appeared in one of the stories. In connection with each story, three sentences were presented as *originals* (i.e., exactly the same sentence as in the presentation phase), one as a *paraphrase* (i.e., different wording with the same meaning at the sentence level), one as a *meaning change* (i.e., different meaning at the sentence level but still compatible with the story plot), and one as a *situation change* (i.e., incompatible with the story plot). Table 2 presents a sample sentence, Appendix D2 contains all the sentences, and Appendix D3 includes the instructions and a practice trial.

For data analysis, acceptance rates, defined as the relative frequencies of positive responses, were calculated for each sentence type; based on them, the question of whether the tripartite model is an appropriate framework for children's text comprehension (*Hypotheses 1a–1c*) was addressed. For predictions regarding the influence of illustrations on children's text comprehension (*Hypotheses 2a–2f*), non-parametric sensitivities (*A'*) adopted from the signal detection theory were calculated (Donaldson, 1992; Stanislaw & Todorov, 1999). *A'* interrelates previously defined hits (*H*) and false alarms (*F*) so that they are represented independently of a participant's response bias (criterion), using the following formula when $H \geq F$.

$$A' = 0.5 + \frac{(H - F)(1 + H - F)}{4H(1 - F)}$$

Values of *A'* range between 0 and 1, with 0.5 representing the chance level. Accordingly, the sensitivity measure for the text surface was calculated with acceptances of original sentences (Y_O) as hits and acceptances of paraphrases (Y_P) as false alarms:

$$A'_{text\,surface} = 0.5 + \frac{(Y_O - Y_P)(1 + Y_O - Y_P)}{4Y_O(1 - Y_P)}$$

For the textbase, combined acceptances of originals and paraphrases ($Y_{O,P}$) were defined as hits and acceptances of meaning changes (Y_M) were defined as false alarms:

$$A'_{textbase} = 0.5 + \frac{(Y_{O,P} - Y_M)(1 + Y_{O,P} - Y_M)}{4Y_{O,P}(1 - Y_M)}$$

Finally, situation model sensitivities were calculated conceiving the acceptances of originals, paraphrases, and meaning changes ($Y_{O,P,M}$) as hits and those of situation changes (Y_S) as false alarms:

$$A'_{situation\ model} = 0.5 + \frac{(Y_{O,P,M} - Y_S)(1 + Y_{O,P,M} - Y_S)}{4Y_{O,P,M}(1 - Y_S)}$$

Notably, *A'* cannot be expressed as a real number if both the hit rate and the false alarm rate equal 0 or 1. In such cases, participants were excluded from hypothesis testing. There is also a statistical interdependence between the sensitivity measures; therefore, a multivariate analysis of variance across all three levels would be inappropriate and separate analyses of variance were performed for each representation level.

Table 2
Original sentences, paraphrases, meaning changes, and situation changes of the first sentence from the story *Die Zeltnacht* ("A Night in the Tent")

Sentence type	German version used in the study	English translation
Original	An einem warmen Sommerabend bauen Tom, Alexa und Maja ein Zelt im Garten auf.	On a warm summer evening, Tom, Alexa, and Maja are putting up a tent in the garden.
Paraphrase	An einem warmen *Abend im Sommer* bauen Tom, Alexa und Maja *im Garten ein Zelt* auf.	On a warm *evening in summer in the garden*, Tom, Alexa, and Maja are putting up a tent.
Meaning Change	An einem *schönen* Sommerabend bauen Tom, Alexa und Maja ein Zelt im Garten auf.	On a *beautiful* summer evening, Tom, Alexa, and Maja are putting up a tent in the garden.
Situation Change	An einem *kühlen* Sommerabend bauen Tom, Alexa und Maja ein Zelt im Garten auf.	On a *cool* summer evening, Tom, Alexa, and Maja are putting up a tent in the garden.

Note. As both meaning and situation changes necessarily imply changes in the text surface, a conservative sentence construction rule was applied to rule out the possibility that rejections of meaning or situation changes could be ascribed to an accurate representation of the text surface. Accordingly, paraphrases contained at least as many changes to the text surface as – and in some cases more than – meaning and situation changes.

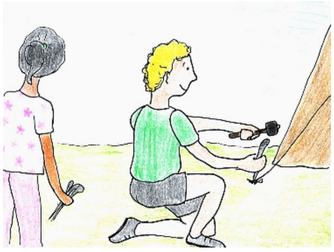
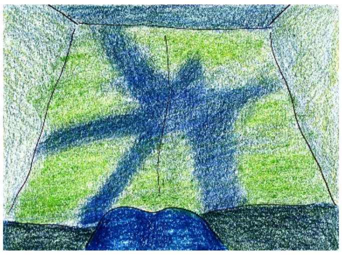
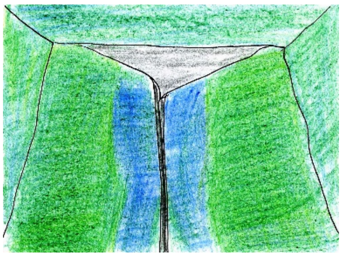
no.	Original sentence (English translation)	Static illustration	Animation
1	On a warm summer evening, Tom, Alexa, and Maja are putting up a tent in the garden.		*Human body:* Tom hammering (left arm moving up and down); Maja approaching
2	They are allowed to sleep outside, so they fetch their sleeping bags and ground pads from the basement.		*Human body:* Tom, Alexa, and Maja walking into the scene (legs moving)
3	When it's getting dark, they snuggl into their sleeping bags and tell each other scary stories.		*Human body:* Maja talking (mouth moving), Alexa's and Tom's eyes blinking
4	Shortly after, Alexa and Maja are snoring loudly; only Tom is lying wide awake in the tent.		*Human body:* Tom's eyes blinking slowly
5	Then he sees a huge shadow on the flysheet, and he crawls deeper into his sleeping bag.		*Other:* Shadow moving
6	Slowly, the zip of the tent opens, and Tom, in relief, hears his mother's whispering voice.		*Other:* Tent zip opening from above

Figure 1. Original sentences, static illustrations, and animation descriptions for the English translation of the story "A Night in the Tent" (for the German version used in the experiments, see Appendix D1).

3.6 Methodology of Experiment 3. Picture verification

Experiment 3 in this thesis involved a picture verification task based on the paradigm introduced by Stanfield and Zwaan (2001) and similar to the one developed by Hauf et al. (2020). The 24 target stories of this experiment consisted of five sentences each, the last of which described a rising or falling target object (see Figure 2 for an English example and Appendix E1 for all stories in their original German versions, including the practice stories). After a half-second interval, a drawing of this target object entered the screen either from above and moved downward or from below and moved upward; see Figure 3 for the outline of an experimental trial. Participants indicated by pressing keys whether the depicted object had occurred in the story. Perceptual simulation was indicated when the response times in matching trials (i.e., identical movement directions in the sentence as in the animated picture to be verified) were shorter than in mismatching trials (movement in opposite directions). This *matching advantage* was quantified by calculating the response time differences between matching and mismatching trials. This calculation included only correct responses to target stories that do not exceed 5 seconds.

In addition, 30 filler stories were included; in six of them, the target object occurred in the story (i.e., a positive answer was required) but was not associated with a vertical movement. In the other 24 stories, the target object was not part of the story, so the correct answer was negative. Under the audiovisual text condition, three static illustrations were presented synchronously with each story; these changed after the first or second and again after the third sentence. The target object was never part of an illustration (see Figure 2 for an example, Appendix E1 for all illustrations, and Appendix E2 for the instructions).

Owing to the fact that the response-time data have a nested two-level structure (participant and item levels), linear-mixed model analyses were run because they enable the computation of fixed and random effects at both levels combined (Richter, 2006; Snijders & Bosker, 2012). The R software (version 3.6.2; R Core Team, 2019) was used for this analysis, including the packages lme4 (version 3.1-1; Bates, Mächler, Bolker, & Walker, 2015) for model construction, lmerTest (version 1.1-21; Kuznetsova, Brockhoff, & Christensen, 2017) for significance testing, and MumIn (version 1.43.15; Barton, 2019) for the determination of explained variance proportions (R^2). Linear-mixed modeling is an iterative procedure that can be conducted via maximum-likelihood estimation; however, as recommended by Bates et al. (2015), a restricted maximum-likelihood method with generalized least squares estimates was employed. Degrees of freedom were estimated via Satterthwaite's method, in line with Kuznetsova et al.'s (2017) recommendation.

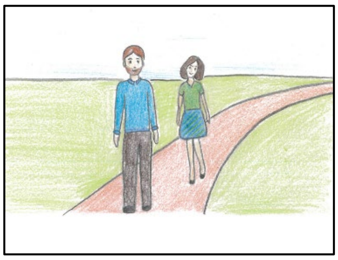
no.	English translation	Picture
1	It's a bright afternoon and Vera and her father are going for a walk on the path.	
2	On their way they see a hot-air balloon standing on the ground.	
3	They wave at the people sitting in the balloon.	
4	Vera is wide-eyed – how much she wants to fly with them!	
5	Suddenly the hot-air balloon rises high into the clouds.	

Figure 2. Sequence of sentences and pictures of an experimental story with an upward object movement.

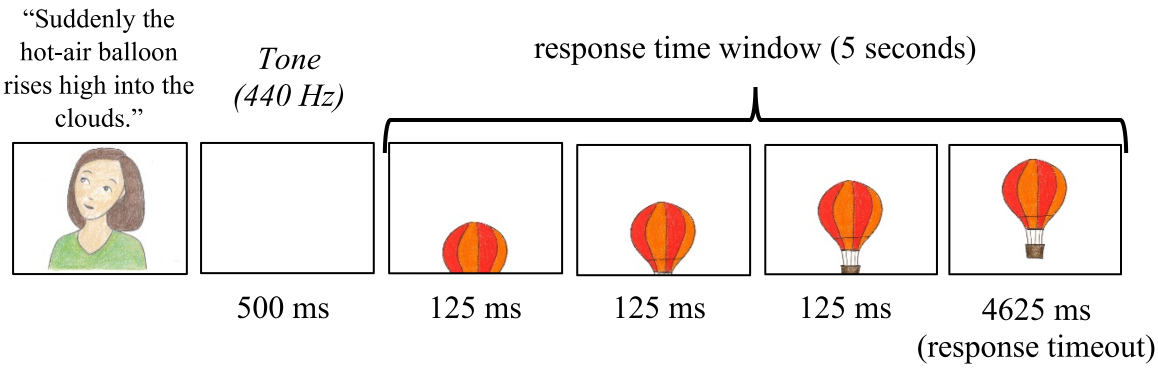


Figure 3. Time course of a matching experimental trial (upward movement) in the audiovisual condition, beginning with the last sentence of the story and ending with the participant's response or response timeout; ms = milliseconds.

4. Experiments

This part of the thesis provides short summaries of the experiments conducted. Experiment 1 addressed children's text surface, textbase, and situation model representations of written narrative text without and with illustrations presented before versus after their corresponding sentences. A submitted manuscript is associated with this experiment (see Appendix A). In Experiment 2, these levels of representation were scrutinized with auditory narrative text among children; auditory text was presented with either dynamic or static illustrations or with no illustrations at all. This experiment corresponds to a research paper published in the *Journal of Cognition and Development* in 2019; the accepted manuscript is presented in Appendix B. In Experiment 3, the perceptual simulation of object movement during the reception of narrative text was studied among children and adults; narrative text was presented aurally either with or without illustrations. A related article appeared in *Discourse Processes* in 2020; the accepted manuscript is presented in Appendix C.

4.1 Experiment 1. Representations of written and illustrated written text

Experiment 1 investigated text surface, textbase, and situation model representations of short written stories in a sample of 146 school children between grades 2 and 6, ranging between 7;9 and 13;0 years of age (mean = 10;5). Three formats of text presentation were varied experimentally within participants: written text without illustrations (referred to as *text-only*), written text with illustrations presented after their corresponding sentences (*text-picture*), and written text with illustrations presented before the sentences (*picture-text*). The hypotheses were that young readers form three levels of representation (*Hypotheses 1a–1c*), yield better text surface representations when the story is illustrated (*Hypothesis 2f*), yield better situation model representations when the story is illustrated (*Hypothesis 2a*), especially when the illustration is processed before its corresponding sentence (*Hypothesis 2b*), and yield poorer textbase representations when the story is illustrated (*Hypothesis 2d*), especially when the illustration is processed after its corresponding sentence (*Hypothesis 2e*). Finally, it was expected that younger participants would profit more from illustrations than older ones regarding the situation model (*Hypothesis 4a*). In the submitted manuscript (Appendix A), reading and picture-viewing times are also reported.

Differences between the acceptance rates for original sentences and paragraphs, paragraphs and meaning changes, and meaning changes and situation changes were each significant, indicating that children do form three levels of representation during reading, supporting *Hypotheses 1a –*

1c. Sensitivities (*A'* measures) did not differ between experimental conditions for the situation model or for the text surface; thus, *Hypotheses 2a, 2b*, and *2f* are rejected. For the textbase, however, a significant effect emerged: textbase sensitivities were lower in the text-picture condition than in the other two conditions, which confirms *Hypothesis 2e*. Thus, processing pictures after their corresponding written text impairs textbase accuracy. There was no significant difference between the text-only and picture-text conditions, so there is no evidence favoring *Hypothesis 2d*, according to which the textbase would be generally impaired when the text is illustrated. Neither sensitivity measure correlated significantly with age; hence, *Hypothesis 4a* was also not supported.

A notable result of this experiment is that illustrations do not appear to improve situation model construction based on written text. This result pattern differs from the one obtained by Wannagat et al. (2018), where higher situation model sensitivities are associated with audiovisual rather than solely auditory text. Taken together, these findings may conform with the modality principle (Low & Sweller, 2014), which states that pictures have a greater impact on auditory than on written text comprehension. Experiment 1 also did not yield results comparable to those of Wannagat et al.'s (2018) study concerning the text surface, where a supporting function of illustrations in auditory stories was observed. This result may nonetheless conform with those authors' explanation that participants can hold more verbatim text information in memory owing to cognitive resources that can be spared for situation model construction in the audiovisual but not in the auditory condition. In this experiment, it is possible that roughly the same amount of cognitive resources were used for situation model construction based on plain and illustrated stories so that no resources could be spared for memorizing the text surface. Alternative explanations of these results are addressed in the General Discussion chapter.

With regard to the textbase, Wannagat et al.'s (2018) result, according to which illustrations impair the semantic processing of auditory text, was also not replicated with written text in this experiment. However, textbase representations of written stories do appear to be impaired by illustrations presented *after* their corresponding verbal text passages. This is in line with the model inspection component of the ITPC framework (Schnotz, 2014): if a situation model is updated on the basis of new information (in this case, pictorial information), an update also applies to the textbase. This update may be functional because it helps readers to articulate their updated situation model verbally; but it nevertheless means that the *original* textbase representation becomes less accurate.

Although the within-participant design of this experiment provided greater statistical power than a between-participant design of the same sample size, two shortcomings emerged that should be addressed. First, participants read only four stories per condition so that they were presented with only four paraphrases, four meaning changes, and four situation changes per condition. This made false alarm rates of 100% quite likely, leading to incalculable *A'* measures with the consequence of serious drop-out rates, especially at the text surface level, where 25.3% of the sample had to be excluded. Second, an analysis of variance that included a group factor indicating which text condition was applied first yielded significant interactions with the experimental within-participant factor for both the situation model and textbase sensitivities. This suggests that the impact of illustrations on textbase and situation model representations is moderated by carryover effects between conditions. These effects can be avoided by applying the experimental variation of presentation formats between rather than within participants. Therefore, Experiment 2 with auditory, audiovisual-static, and audiovisual-dynamic text conditions followed a between-participant design.

4.2 Experiment 2: Representations of auditory, audiovisual-static, and audiovisual-dynamic text

In this experiment, text surface, textbase, and situation model representations of auditory stories were examined in a sample of 36 grade 2 (mean age = 7;9, *SD* = 0;6), 36 grade 4 (mean age = 9;10, *SD* = 0;6), and 36 grade 6 (mean age = 11;9 *SD* = 0;6) students. Three text versions were varied experimentally between participants: one auditory without illustrations (*auditory*), one audiovisual with static illustrations (*AVS*), and one audiovisual with dynamic illustrations (*AVD*). The hypotheses were that listeners build text surface, textbase, and situation model representations (*Hypotheses 1a–1c*), yield better text surface representations in AVD and AVS than in auditory stories (*Hypothesis 2f*), yield better situation model representations in AVS and AVD than in auditory stories (*Hypothesis 2a*), yield better situation model representations in AVD than in AVS stories (*Hypothesis 2c*), and yield poorer textbase representations in AVD and AVS than in auditory stories (*Hypothesis 2d*). It was further hypothesized that younger participants would yield more accurate situation models in audiovisual (both AVD and AVS) than in solely auditory stories than their older counterparts (*Hypothesis 4a*).

Again, acceptance rates differed significantly between original sentences, paraphrases, meaning changes, and situation changes; this confirms *Hypothesis 1* as a whole, suggesting that children form three levels of representation when listening to narrative text. Sensitivity measures (*A'*) differed significantly across conditions for the text surface, indicating that children in the AVS condition outperformed those in the auditory condition; however, there was no difference between AVD and auditory or between AVD and AVS conditions, which partly confirms *Hypothesis 2f*. Textbase *A'* measures were not affected by experimental conditions, so *Hypothesis 2d* is rejected. For the situation model, a significant effect emerged indicating higher sensitivities in the AVS and AVD conditions than in the auditory condition, thus corroborating *Hypothesis 2a*. However, there was no difference between AVD and AVS texts, yielding no evidence for *Hypothesis 2c*, according to which dynamic illustrations would improve situation model construction in particular. The assumption that younger children's situation models would benefit from illustrations more than that of older ones (*Hypothesis 4a*) was, again, not supported.

This experiment can be seen as a successful replication of earlier results (Wannagat et al., 2018), insofar as illustrations help children construct more accurate situation models; it can also be considered an extension of those earlier efforts by suggesting that both static and dynamic illustrations help listeners construct a coherent representation of the state of affairs addressed in the text. Likewise, regarding the text surface, there is evidence that the facilitating effect of illustrations on the situation model frees up cognitive resources that can be used for memorizing the text surface. However, this appears to pertain only to AVS and not to AVD text. One explanation may be that animations demand additional cognitive load (Ayres & Paas, 2007) and may have counteracted the otherwise positive effect of illustrations on text surface representations. Textbase sensitivities were not affected by illustrations in this experiment; this is not in line the assumption derived from the ITPC (Schnotz, 2014), which states that obtaining a situation model on the depictive path would render the descriptive path less relevant (Wannagat et al., 2018). Nonetheless, the positive effect of illustrations on the situation model – at least when the text is presented orally – conforms with the ITPC's central claim that text-picture units activate two different processing paths leading to the situation model: construction on the descriptive path and analog structure mapping on the depictive path. Both paths appear to result in practically the same situation model; in a consequence, that model may be more accurate than a situation model obtained solely through construction on the descriptive path.

In terms of the embodied approach to the situation model (Glenberg & Robertson, 2000; Zwaan, 2014; Zwaan & Radvansky, 1998), the results are mixed. On one hand, there was virtually no difference between AVD and AVS text regarding situation model sensitivities, so motor resonance cannot be considered relevant for text comprehension on the basis of the results reported here – or at least, not as more relevant than the general proposition that perceptual sensorimotor processes are functionally related to situation models based on illustrated text (Wannagat et al., 2018). On the other, this latter expectation is at least not contradicted by the results of this experiment, although the findings from Experiment 1 with written text do not favor it. According to this expectation, illustrations initiate and support perceptual simulation, and perceptual simulation helps the text recipient obtain an accurate situation model. The first part of this proposition was addressed in Experiment 3.

4.3 Experiment 3: Perceptual simulation in auditory and audiovisual text

Experiment 3 was performed to examine the perceptual simulation of vertical object movement during text comprehension in samples of children and adults; it was carried out because there have only been few attempts to address perceptual simulation among children, especially as to the dynamic features of the situation described in the text. Moreover, most previous studies with children (de Koning et al., 2017; Engelen et al., 2011; Hauf et al., 2020) have been conducted with sentences, so this experiment constituted an innovation insofar as whole narrative texts were used as text stimuli. In addition, the effect of illustrations on perceptual simulation was under investigation; to my knowledge, this has been studied only once (Hauf, 2016) with regard to static object features. A picture verification method adapted from Hauf (2016; Hauf et al., 2020) was used, holding that perceptual simulation is indicated by shorter response times in matching than in mismatching trials (*matching advantage*). It was predicted that participants would perceptually simulate vertical object movements (i.e., exhibit a matching advantage; *Hypothesis 3a*) and that a more pronounced perceptual simulation (i.e., higher matching advantage) would occur among participants in the audiovisual than in the solely auditory condition (*Hypothesis 3b*). As a specification of the perceptual boundedness/perceptual support hypothesis (Ruggeri & Katsikopoulos, 2013; Springer, 2001; Unsöld & Nieding, 2009) and the suggestion that embodied cognition is the motor through which illustrations provide perceptual support (cf. Pexman, 2019), it was also expected that younger participants would display a more pronounced matching advantage than older ones (*Hypothesis 4b*), especially when exposed to audiovisual text (*Hypothesis 4c*).

Preschoolers (mean age = 6;0, *SD* = 0;4, *N* = 46), second graders (mean age = 8;2, *SD* = 0;6, *N* = 48), fourth graders (mean age = 10;3, *SD* = 0;5, *N* = 48), and young adults (mean age = 23;0, *SD* = 2;10, *N* = 48) took part in this experiment. Matching and mismatching of vertical object movements between the story and the target picture varied within participants, whereas the presentation format of the stories (auditory vs. audiovisual) varied between them.

A matching advantage occurred, indicating that adults and children perceptually simulate vertical object movements, so *Hypothesis 3a* is confirmed. This matching advantage did not vary across presentation formats, so there is no evidence that perceptual simulation is affected by the presence or absence of illustrations; *Hypothesis 3b* is thus rejected. Age groups exhibited significantly different response times, with older participants performing more rapidly than younger ones; this likely reflects general differences related to the development of processing speed during childhood (Kail & Salthouse, 1994). However, there was no significant variation in the matching advantage across age groups, which did not conform to the assumption that younger children would exhibit a greater matching advantage than older ones, so *Hypothesis 4b* is rejected. There was also no significant variation in the matching advantage across age groups or presentation formats (i.e., three-way interaction); therefore, *Hypothesis 4c* is rejected as well. Against expectations, there were significantly faster response times in the auditory than in the audiovisual condition; this difference could be associated with differences regarding the response cue strength of pictures. In the auditory condition, the only pictures that ever appeared showed the objects to be verified, so each picture was associated with a due response. In the audiovisual condition, by contrast, three pictures served as illustrations of the auditory text, so only one of four pictures was associated with a due response. Thus, participants in the audiovisual condition may have taken longer to realize that they had to provide a response when the target picture was shown.

The results indicate that both children aged 6 years and older as well as adults perceptually simulate vertical object movements during narrative text comprehension, and that this simulation can be considered a fairly automatic and effortless process. An earlier study provided evidence for the perceptual simulation of vertical object movements during sentence comprehension (Hauf et al., 2020); Experiment 3 replicates and extends this finding to story comprehension. As in their experiment, children and adults exhibited a matching advantage regardless of age, suggesting that perceptual simulation may be fully developed at school entry (see also de Koning et al., 2017; Engelen et al., 2011) or even before that point (Fecica & O'Neill, 2010). Unexpectedly, however, the result pattern reported here does not allow for the conclusion that illustrations improve the perceptual simulation process during text comprehension. The implications of this absence of evidence for the functionality of perceptual simulation regarding the situation model are discussed in the next chapter.

5. General Discussion

This part of the thesis integrates the results from the experiments described in the previous chapter with the theoretical and empirical research body on text comprehension outlined earlier in this synopsis. More specifically, it addresses whether children form three levels of representation during text comprehension, how illustrations affect these levels of representation, and how perceptual simulation and motor resonance are related to the comprehension of illustrated text among children. Furthermore, implications for the development of narrative text comprehension within this age range are addressed. Finally, an explanation model for the insights reported here is provided before the methodological limitations of this work and suggestions for future research are presented.

5.1 Text surface, textbase, and situation model

First, the results obtained here suggest that children build representations of the text surface and textbase and construct a situation model while reading (Experiment 1) or listening to (Experiment 2) narrative text. This conforms with earlier studies in which Schmalhofer and Glavanov's (1986) sentence recognition method has been applied to narrative text among children (Nieding, 2006; Wannagat et al., 2017, 2018). This result also demonstrates the construct validity of the sentence recognition method for the sample of participants and stories used in this thesis.

From a theoretical point of view, it is still not clear to what extent *representations* of text reflect its *comprehension*. As already noted, the text surface, which has usually been assessed via literal information questions in studies of reading comprehension, is not substantially related to comprehension ability (Cain & Oakhill, 1999; Oakhill et al., 2003); its necessity for research on text comprehension may therefore be questioned. Notwithstanding this point, it can be argued that the text surface is important from both theoretical and practical points of view. Theoretical perspectives on text comprehension as a process imply that the text surface serves as a starting point of comprehension; this applies to reading comprehension models (e.g., Perfetti et al., 2005; Verhoeven & Perfetti, 2008) and to the ITPC (Schnotz, 2014) which is addressed in Section 5.2. Moreover, the discriminatory power of the textbase against both the text surface (meaning changes vs. paraphrases) and the situation model (meaning changes vs. situation changes) should be reflected. Of course, the empirical results of Experiments 1 and 2 confirm that paraphrases are less dramatic alterations of the text than meaning changes, whereas situation changes must be deeper than meaning changes within the research described here. However, when constructing a sentence recognition task, this needs to be clarified beforehand. As to meaning versus situation changes, Fletcher and Chrysler (1990) have demonstrated in a series of experiments that these two sentence types are related to distinct levels of comprehension; therefore, it should suffice to know whether the modified sentence is consistent with the story plot. However, the case of meaning changes versus paraphrases is less clear, especially when synonyms are used in paraphrases. Perfetti et al. (2008) have demonstrated that the context determines whether certain words can be used as synonyms (e.g., *spilling* vs. *emptying* a glass of orange juice), so it may be relatively difficult to draw a bright line between paraphrases and meaning changes.

5.2 Multimedia comprehension and narrative text

Theories that refer to the comprehension of text-picture units such as the multimedia principle (Butcher, 2014; Mayer, 2009), the modality principle (Low & Sweller, 2014), and the ITPC (Schnotz, 2014) originate from research on learning from expository text that does or does not include graphical information (e.g., Butcher, 2006; Cuevas et al., 2002; Eitel et al., 2013; Glenberg & Langston, 1992; Moreno & Mayer, 1999; Mousavi et al., 1995), whereas similar research with narrative text (e.g., Guttmann et al., 1977; O'Keefe & Solman, 1987; Pezdek et al., 1984) often lacks such a theoretical underpinning. Therefore, applying these theories is an obvious step toward obtaining deeper insights into the comprehension of illustrated narrative text. By and large, the results reported in this thesis support the conclusion that applying these theories in connection with narrative text is appropriate. Both the multimedia and modality principles can be subsumed under the ITPC; therefore, the argumentation below refers mainly to the ITPC.

5.2.1 Auditory versus written text

The ITPC holds that the first stage of the comprehension process is perceptual surface structure processing (Schnotz, 2014) and depends on whether verbal and pictorial text input enters the mind from two sensory registers (auditory-visual) or only one (visual-visual; this is what the modality principle refers to). Accordingly, illustrated text that enters from two registers should lead to more accurate representations than illustrated text that enters from only one source. Although none of the present experiments investigated this assumption directly, they do provide evidence for it when considered in combination: the situation model improves when illustrations are associated with auditory text (Experiment 2; see also Wannagat et al., 2018) but not when they are presented with written text (Experiment 1), when both are compared with a non-illustrated version of auditory or written text. The same trend appears to apply to the text surface, at least when static illustrations are used (see also Wannagat et al., 2018).

However, given that written text and illustrations were presented asynchronously, there are at least two alternative explanations for not finding an advantage of illustrated over solely written text regarding the situation model. First, earlier studies have revealed a positive effect when illustrations were presented simultaneously with their corresponding written text (Gambrell & Jawitz, 1993; Pike et al., 2010) or have found a greater advantage of simultaneous text-picture units than of sequential ones (O'Keefe & Solman, 1987). Therefore, it appears likely that concurrent units of verbal text and pictures have unique features that facilitate situation model construction and are not shared by sequential ones; this is addressed more thoroughly in the next subsection. Second, in Experiment 1, sentences were presented in alternation with their corresponding pictures, whereas O'Keefe and Solman presented the whole set of pictures before or after the complete written text. The approach taken in Experiment 1 here meant that illustrations interrupted the reading flow and thus may have caused their potential benefits to disappear, whereas there was no such interruption in O'Keefe and Solman's experiments. In sum, it can be stated that the evidence for the modality principle based on the present research is fairly weak; therefore, direct comparisons between written and auditory narrative text with illustrations with regard to tripartite representations are recommended.

Another possible explanation for the absence of positive effects of illustrations on situation models based on written text may stem from differences regarding expertise in reading versus listening among children in their early school years; these children may not benefit from illustrations associated with written text because they are, as novice readers, too busy with word decoding. This view accords with early research on reading comprehension suggesting that listening comprehension surpasses reading comprehension at approximately grade 7 (Sticht, Beck, Hauke, Kleiman, & James, 1974), which is still later in the schooling process than the students participating in the present research. However, newer research suggests that reading and listening comprehension of narrative text are at the same level by grade 4 (Diakidoy, Stylianou, Karefillidou, & Papageorgiou, 2005; Wannagat et al., 2017); this would necessarily imply that situation models improve with age, which was not the case in the present study. An opposing but also plausible argument is that the more difficulties children have with reading, the more they would benefit from additional information obtained from pictures. This would be in line with the scaffolding hypothesis (cf. Eitel & Scheiter, 2015) which is discussed in the next subsection.

5.2.2 Situation model construction and analog structure mapping

The second step of the ITPC, semantic deep structure processing, is accomplished on two paths, one descriptive and one depictive (Schnotz, 2014; Schnotz & Bannert, 2003). On the descriptive path, a textbase representation is formed based on the text surface representation, and a situation model is constructed based on the textbase and previous knowledge (van Dijk & Kintsch, 1983). On the depictive path, a situation model can be obtained directly via analog structure mapping based on a surface representation of the picture provided (Gentner, 1989). According to this reasoning, it was expected that the situation model would improve when illustrations were added to verbal text. The results reported here suggest that this is true, at least for auditory text (Experiment 2), and it can be said that there is additional evidence in favor of this assumption, given that Wannagat et al. (2018) obtained similar results using a similar method. Moreover, this result corroborates earlier studies where a greater accuracy in tasks related to the situation model such as inference tasks has been associated with audiovisual text than with solely auditory text (e.g., Beagles-Roos & Gat, 1983; Gibbons et al., 1986; Hayes et al., 1986).

However, this association is less clear in the case of written text. The failure to find an advantage of illustrated over non-illustrated written text might have its origin in the early processing step, according to which it is better to have two sensory channels involved instead of one, but there is an alternative explanation, according to which a simultaneous presentation of written text and illustrations would be superior to a sequential one like that implemented in Experiment 1 (O'Keefe & Solman, 1987). Simultaneous text-picture units may yield more accurate situation models because they enable readers to process text and illustration in an iterative manner; future studies could investigate this assumption via eye-tracking indicators such as the number of saccades between text and illustration.

Based on the analog structure mapping hypothesis, it was further predicted that accuracy on the situation model would be higher when illustrations were presented before rather than after their corresponding written sentences. Accordingly, a picture could serve as a *scaffold* for the subsequent process of situation model construction (Eitel & Scheiter, 2015; Schnotz & Bannert, 2003). The results in Experiment 1 do not support this scaffolding assumption, indicating that it might not be applicable to the domain of narrative text comprehension. This view, however, is far from being decisive. Indeed, Eitel and Scheiter (2015) have reported that the scaffolding effect depends on the complexity of each stimulus and that the opposite effect (an easy verbal text scaffolding a more complex picture) may also be plausible (cf. Ainsworth, 2006). In the present research, the narrative text was relatively simple (at least for the older participants) when compared with common expository text. At the same time, it featured a relatively high number of pictures, with each picture accompanying no more than 15 words on average. This suggests that the sentences and illustrations used here may not differ substantially in their complexity and that, consequently, scaffolding could not take place. Future research should use more complex stories with lower picture-per-word rates to investigate the scaffolding hypothesis in greater depth.

At the end of the previous subsection, it was argued that *novice* readers' situation models could benefit from the scaffolding function of an illustration. In this case, one would not expect a general advantage of the picture-text over the other two presentation formats (text-only and text-picture) in Experiment 1, but a particular advantage of this kind among students below grade 4, but not older ones (Diakidoy et al., 2005). This is also not supported by the present data, for which there are at least two alternative accounts. On one hand, all participating children may be comparative novices in reading, in line with Sticht et al.'s (1974) integrative work on early reading comprehension research, whereas there is no scaffolding function of illustrations in the present study. On the other, it is possible that even the second graders – who may still struggle with word decoding – are familiar enough with the specific text genre used in the present research (i.e., short pieces of narrative text referring to commonplace events in children's lives) to count as *experts* in this context. This would imply that the illustrations presented in Experiment 1 may have had a scaffolding function, but the participating children had no need for scaffolding. To sum up, the present research suggests that reading expertise is of limited relevance for text comprehension.

5.2.3 Textbase and model inspection

If there is a direct way to obtain a situation model on the depictive path (analog structure mapping), it can be argued that the descriptive path (situation model construction) is rendered less relevant. Based on this reasoning and on Wannagat et al.'s (2018) results, it was hypothesized that illustrations would yield less accurate textbase representations than would occur in their absence. This result did not emerge in the experiments reported here, suggesting that text recipients use their textbase representations to construct situation models regardless of whether a story is illustrated. Thus, structure mapping processes on the depictive path may supplement the construction process, but they do not replace it.

Another prediction derived from the ITPC was that textbase representations are affected by model inspection processes; these were assumed to originate from illustrations that are presented after their corresponding sentences and to result in poorer textbase representations in the text-picture than in the picture-text condition of Experiment 1. The results confirm this hypothesis, indicating that model inspection occurs during narrative text comprehension and that it affects not only the situation model but also the underlying semantic representation of verbal text. In accordance with Schnotz (2014), it can be suggested that recipients overwrite the textbase after changes to the situation model in order to prepare verbal utterances related to the stories to which they have been exposed.

5.3 Embodied accounts of the situation model

Researchers in the field of embodied cognition (e.g., Barsalou, 1999; Glenberg & Robertson, 2000; Zwaan, 1999, 2014) have argued that the situation model is a multimodal representation based on neural images of an individual's experience of the outer world. Aspects of the situation described in a text are simulated via the individual's sensory (perceptual simulation; Kaschak et al., 2005; Stanfield & Zwaan, 2001) and motor systems (motor resonance; Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006). It has been suggested that the analog structure mapping process claimed by the ITPC is also not merely symbolic, but can rather be associated with processes such as perceptual simulation and motor resonance (Wannagat et al., 2018). There are two necessary premises of this assumption. First, of course, the analog structure mapping hypothesis must be confirmed; this means that illustrated text should yield more accurate representations of the situation model than non-illustrated text. This appears to be the case, at least in auditory text (Experiment 2). Second, if there is an association between perceptual simulation and analog structure mapping, the former should be more pronounced when the latter is possible; namely, when illustrations are present.

Experiment 3 was designed to examine this assumption as to perceptual simulation. The results suggest that, in line with earlier studies of perceptual simulation (Engelen et al., 2011; Hauf, 2016; Hauf et al., 2020), children aged 5 to 11 years and adults simulate vertical object movements that occur in narrative text. However, perceptual simulation does not appear to be affected by the presence or absence of illustrations, so there is no evidence for the second premise outlined above: that perceptual simulation would be more pronounced when the text is illustrated. Therefore, the present results do not yield direct evidence for the claim that there is a functional relationship between perceptual simulation and the situation model (i.e., that perceptual simulation helps children with text comprehension). Regarding motor resonance in particular, the first of these premises – that *dynamic* illustrations would evoke superior situation models when compared with static ones – had not previously been investigated, so this was done in Experiment 2. The results of that experiment do not support this premise; thus, in the case of motor resonance, there is no evidence that it is associated with situation models based on narrative text. Strictly speaking, failures to reject the null hypothesis do not allow for the conclusion that there is no effect, so there should be replication studies to substantiate these results with greater statistical power. Moreover, there are alternative explanations for both results that are addressed later on in the limitations section.

Nevertheless, there are also reasons to believe that processes related to embodied cognition play only a minor role in situation model construction when one considers the recent literature in this field. In early publications (e.g., Glenberg & Kaschak, 2002; Kaschak et al., 2005; Taylor & Zwaan, 2009; Zwaan, 1999), it was claimed that perceptual simulation and motor resonance are by necessity functionally linked to text comprehension; this position refers to a *strong embodiment view* (Mahon, 2015), according to which all linguistic concepts are represented in modality-specific sensorimotor formats. More recently, however, researchers on this topic (e.g., Mahon & Caramazza, 2008; Zwaan, 2014) have advocated for an integration of embodied and symbolic approaches, suggesting that sensorimotor processes are somehow related to the formation and/or the reactivation of linguistic concepts and thus *can* be functionally linked to text comprehension. Therefore, one of the most prominent challenges that research on embodied cognition should confront is whether perceptual simulation and motor resonance are necessary for – or at least supportive of – text comprehension or whether they are mere by-products of text reception (Kaup, de la Vega, Strozyk, & Dudschig, 2015; Ostarek & Huettig, 2019). Strozyk, Dudschig, and Kaup (2019) examined this issue by applying a dual-task paradigm to a lexical decision task of hand- and foot-related words. No difference emerged between dual-task (hand or foot tapping during lexical decision) and single-task (lexical decision only) conditions, indicating that sensorimotor processes may be only a by-product of language comprehension, at least at the word level.

Another recent study (Ostarek, Joosen, Ishag, de Nijs, & Huettig, 2019) examined the influence of visual distractors on the perceptual simulation of object shapes during sentence comprehension. The authors used Zwaan et al.'s (2002) picture verification method, in which participants were exposed to a sentence (e.g., "The ranger saw the eagle in the sky") and then asked whether a subsequently presented picture referred to an object that was part of the sentence; in critical trials, this picture either matched (e.g., eagle with spread wings) or mismatched (eagle with folded wings) the object form implied by the sentence. The participants in Ostarek et al.'s (2019) study were exposed to visual distractors while listening to sentences so that the distractors would have the potential to interrupt perceptual simulation (cf. Kaschak et al., 2005). Two types of distractors were included: *visual noise* consisting of unrecognizable multicolor shapes and *semantic noise* depicting commonplace objects that were not mentioned in any of the sentences. In line with the assumption that the simulation is a low-level visual process, it had been expected that visual but not semantic noise would inhibit perceptual simulation (i.e., a matching advantage had been expected in semantic noise, but not in visual noise conditions). However, an opposite pattern of results emerged: a matching advantage occurred when participants were exposed to visual, but not semantic noise. Ostarek et al. (2019) argue that the matching advantage is not related to basic visual but rather to semantic processing and that, as a consequence, picture verification might not be an appropriate measure of perceptual simulation. They further suggest that insights based on the picture verification paradigm do not uniquely support a strong embodiment view; they are also compatible with an integrated account (Mahon, 2015; Mahon & Caramazza, 2008), according to which amodal concepts are connected with the sensory and motor systems, but in a way that sensory states do not have a substantial impact on the amodal processing of sentences or even longer text units.

Therefore, the present results might be more reconcilable with such an integrated theory than with a strong embodiment view because they do not support the assumption that the sensorimotor re-experiencing of events is directly linked to the improvement of situation model construction in illustrated compared with non-illustrated text. As Ostarek and Huettig (2019) have already suggested, future investigations should specify which types of processes underlie the matching advantage in picture verification tasks and develop more decisive paradigms to address the functional role of perceptual simulation and motor resonance in text comprehension. As a final remark on this topic, I propose that it might be more fruitful to further examine and apply integrated theories of modal and amodal representations rather than strong embodied approaches to text comprehension research. In line with Pexman's (2019) argumentation, this may also pertain to developmental issues regarding this topic that are discussed in the next section.

5.4 Perceptual boundedness and perceptual support

Understanding how text comprehension – especially the comprehension of multimedia text – develops during childhood was one of the main goals of the present research. In accordance with Springer (2001), it was expected that during text comprehension, younger children would overrepresent perceptual information provided in the text (perceptual boundedness) and use this information to foster the development of comprehension-related skills such as conceptual knowledge (perceptual support). If these two phenomena were relevant in the domain of narrative text comprehension, both would predict that the beneficial effect of illustrations on the situation model would be more pronounced among younger than among older children. The present results, however, do not favor this hypothesis, regardless of whether written (Experiment 1) or auditory text (Experiment 2) was presented.

This outcome is consistent with two earlier studies using similar methods (Wannagat et al., 2017, 2018). Assuming that the combined results of these four experiments conducted by our research group allow for the conclusion that there are no age-related differences concerning the impact of illustrations on text comprehension, I propose three alternative explanations related to the perceptual support framework. First, perceptual support, which can be associated with concept development (Springer, 2001) and word learning (Hald et al., 2015), may not apply to the comprehension of longer and more complex language units such as sentences or even stories. Second, it is possible that sentence recognition tasks like those used in our experiments are insensitive toward developmental changes in perceptual support. One plausible specification of this account may be that children in their early school years exhibit perceptual support only with measurements that explicitly demand inference generation, where at least two studies (Pike et al., 2010; Ruggeri & Katsikopoulos, 2013) have found age-related differences for which perceptual support can account. In a similar manner, it can be argued that the stories used in our experiments were too readily comprehensible to find any possible effect related to perceptual support. Third, it may be that perceptual support plays only a marginal role in children aged 6 years and older but is more profoundly associated with younger children. This is consistent with the majority of empirical findings that Springer (2001) refers to in his theoretical work; most suggest that perceptual support plays a central role among preschoolers aged 4 or 5 but is less crucial among primary school children who are 8 or 9 (see, e.g., Gentner & Toupin, 1986).

Another prediction was made based on the perceptual boundedness theory; namely, that the perceptual simulation of vertical movement would be more pronounced among 6 year-old children than among older children or adults. The data obtained in Experiment 3 do not support this assumption, indicating that perceptual boundedness plays no substantial role in the present research. There is evidence that the perceptual simulation of dynamic features of the situation described in narrative text is already developed among preschoolers (e.g., Fecica & O'Neill, 2010), so there may not be any developmental change among children aged 6 and older. From the argument that perceptual simulation is a by-product of text reception that does not support comprehension (cf. Strozyk et al., 2019), it follows that not simulating comes with a more parsimonious application of cognitive resources. On that note Hauf's (2016) results regarding the perceptual simulation of object shapes comply with the perceptual boundedness view because 6 year-old preschoolers – but not older children or adults – have demonstrated a matching advantage. Older children and adults suppress unnecessary perceptual information, whereas preschoolers still fail to do so. The present results regarding the perceptual simulation of vertical movement, indicating that all participants (including adults) simulate vertical movement (see also Hauf et al., 2020) do not corroborate this earlier finding. One explanation may be that shape is a more complex feature than vertical movement so that no significant amount of cognitive resources can be spared by suppressing the simulation of vertical movement.

Based on Pexman's (2019) reasoning that embodied cognition plays a role in younger, but not older children's acquisition of language, it was further argued that perceptual simulation could also be functionally related to perceptual support in the domain of text comprehension. Accordingly, illustrations were expected (a) to be processed via perceptual simulation and (b) to provide younger children with perceptual support. Therefore, and in accordance with earlier findings regarding children's reception of narrative text (Hauf, 2016; Unsöld & Nieding, 2009), it was also hypothesized that 6 year-old but not older participants would exhibit a more pronounced perceptual simulation when illustrations were present than when they were absent. However, the results reported here also fail to support this hypothesis, indicating that perceptual simulation may be independent of age, regardless of whether auditory or audiovisual text is presented; thus, it might be posited that perceptual simulation is unrelated to perceptual boundedness or perceptual support during childhood. In the light of recent research on the role of embodied cognition during cognitive development (e.g., Pexman, 2019), this seems quite unlikely; for this reason, I propose two more specific accounts of the results. On one hand, perceptual support may play no substantial role in the development of text comprehension, as opposed to cognitive processes that develop earlier in childhood, especially conceptual development (Pexman, 2019; Springer, 2001) and word learning (e.g., Hald et al., 2015). On the other, the perceptual simulation of shape, which has been addressed in Hauf's (2016) research, may be more complex and thus demand more effort than the perceptual simulation of vertical movement examined here. This may imply that, in the present research, not simulating does not save cognitive resources even when it is not beneficial for text comprehension.

Again, the conclusions drawn here are based on results that support the null hypothesis and should therefore be regarded with some caution, especially because they involve interactions of the age variable with experimental factors. It is less likely in cross-sectional designs to find an interaction between an experimental factor and age than one of the involved main effects. This applies to Experiment 3 in particular, where the within-participant matching advantage is rather small ($d = 0.21$) but still statistically significant. Power analyses using GPower (Faul, Erdfelder, Lang, & Buchner, 2007) reveal that a matching x age group interaction (corresponding to *Hypothesis 4b*) of the same effect size would be associated with a power of .66, whereas for a three-way interaction (matching x age group x presentation format, corresponding to *Hypothesis 4c*), the power would be only .50. For that reason, more developmental research on the comprehension of illustrated and non-illustrated text should be carried out to provide more decisive evidence for the development of text comprehension.

5.5 Explanatory model and implications

This thesis refers to a bundle of experiments that addressed the comprehension of illustrated versus non-illustrated narrative text among children in their early school years. The results confirm that, in line with the tripartite model provided by van Dijk and Kintsch (1983), children form representations of the text surface and textbase and construct a situation model during text reception. In addition, the results are consistent with predictions based on Schnotz's (2014) ITPC, suggesting that the ITPC, which originated in the field of learning from expository multimedia text, can also be applied to narrative text. Specifically, and in line with earlier results (Wannagat et al., 2018), children's situation models benefit from illustrations because they allow for obtaining a situation model on an additional depictive path via analog structure mapping (Gentner, 1989; Schnotz & Bannert, 2003), at least when the verbal text is presented orally. Moreover, this research provides evidence for the model inspection hypothesis; namely that when a constructed situation model is updated based on additional (here, depictive) information, the textbase is adapted to this new situation model so that the representation of the original textbase is weakened. Finally, it appears that text surface representations are more accurate in the same text formats that also improve the situation model, suggesting that in these formats, situation model construction demands fewer cognitive resources, leaving more to be used to remember the exact wording of the text. Figure 4 depicts an explanatory model for both previous results and those presented here concerning the comprehension of illustrated narrative text among children in their early school years.

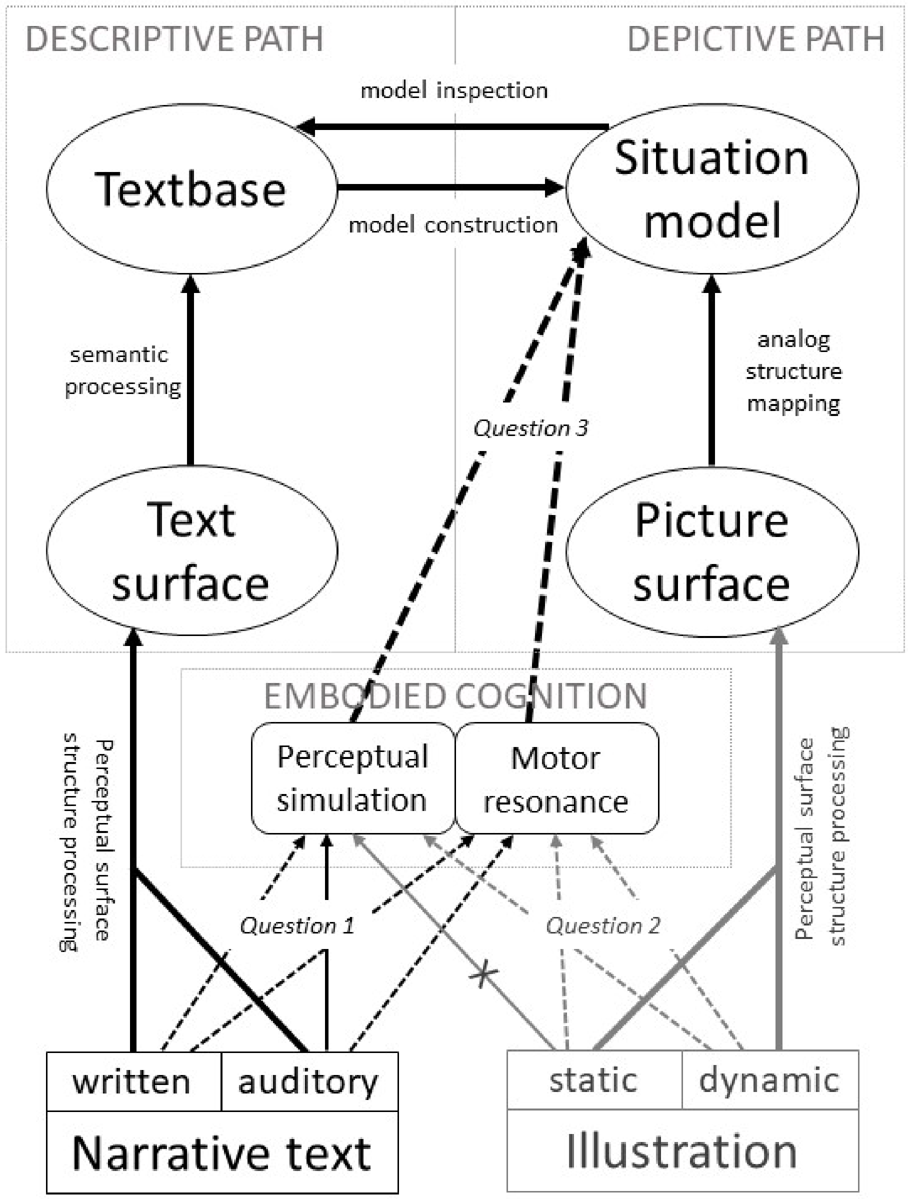


Figure 4. Explanatory model integrating the evidence from previous and present results regarding children's comprehension of illustrated narrative text and embodied cognition processes during text reception (cf. Schnotz, 2014). Solid lines refer to evidence favoring the depicted relationship, while solid lines with an *X* indicate that results presented here do not support the assumed relationship. Dashed lines indicate that the depicted relationship has not yet been addressed among children. *Question 1.* Do children exhibit embodied cognition during text reception? *Question 2.* Do illustrations have an influence on embodied cognition during children's text reception? *Question 3.* Do embodied cognition processes yield a functional relationship with narrative text comprehension (i.e., the situation model) among children?

As to embodied cognition processes (Barsalou, 1999; Glenberg & Robertson, 2000; Taylor & Zwaan, 2009), there are mixed results. On one hand, the present research adds evidence to previous findings suggesting that children perceptually simulate features of a being situation described verbally (e.g., Engelen et al., 2011; Hauf et al., 2020) and extends these earlier findings by indicating that perceptual simulation also takes place when stories rather than sentences are presented. On the other, there is no direct evidence that the presence of illustrations has an impact on perceptual simulation, so it seems unlikely that perceptual simulation contributes to the advantage of illustrated over non-illustrated text regarding the situation model. As Figure 4 shows, motor resonance processes in the course of narrative text comprehension among children deserve more attention in future research. It should be clarified whether motor resonance occurs during children's text reception just as perceptual simulation does; to my knowledge, this has only been done with sentence reception among adults (Glenberg & Kaschak, 2002; Zwaan et al., 2010; Zwaan & Taylor, 2006). Conversely, whether illustrations – especially dynamic ones – affect motor resonance has not yet been adequately examined. A further remaining question is whether children in their early school years yield embodied cognition processes during *reading*; arguably, this question is not trivial given that novice readers have to invest more effort into decoding (Diakidoy et al., 2005; Perfetti et al., 2005; Sticht et al., 1974), which might leave fewer cognitive resources for perceptual simulation and motor resonance. Finally, and importantly, I agree with Ostarek and Huettig (2019) and Kaup et al. (2015) that further investigations should address the extent to which embodied cognition is functional to text comprehension (i.e., the situation model) among adults and, especially, in the course of childhood development (see also Pexman, 2019).

Regarding developmental perspectives on the comprehension of illustrated text, it appears that Springer's (2001) perceptual boundedness and perceptual support hypotheses cannot account for the present results. As argued in Chapter 5.2 in the context of written text, children within the age range examined here do not benefit from perceptual support because they are either reading novices that are too occupied with decoding or experts in the particular text genre to which they were exposed, which means that they are not dependent on perceptual support. The present findings based on auditory text suggest that children, regardless of age, benefit from illustrations. It seems *prima facie* that this favors the first explanation: that children through grade 6 benefit from perceptual support provided by pictures. On closer inspection, however, this *novice argument* has considerable shortcomings so that the alternative *expert argument* appears more plausible. First, the beneficial effect of illustrations affects listening comprehension, at which children in grade 2 are regarded as experts (Sticht et al., 1974). Second, research in the field of multimedia comprehension has repeatedly revealed that adults also benefit from pictures (e.g., Butcher, 2014; Mayer, 2009; Schnotz, 2014), so the multimedia principle (or the ITPC) can account for the present results at least as well as the perceptual support hypothesis. Finally, most previous results that can be explained by perceptual boundedness or perceptual support (e.g., Gentner & Toupin, 1986; Hauf, 2016; Isbell et al., 2004; Pike et al., 2010; Ruggeri & Katsikopoulos, 2013; Unsöld & Nieding, 2009) are obtained with children aged 8 years and younger, so it seems implausible that they would still play a role among 12 year-old children in the present work, especially if one takes into account that the experiments reported here used stories events from children's daily lives. As it appears unlikely that the stories used here included novel concepts that would require perceptual support (cf. Pexman, 2019; Springer, 2001), I suggest that, although pictorial information still improves children's text comprehension owing to the availability of a depictive processing path, the widespread assumption that children are closely tied to illustrations is overestimated in some way, at least in the field of narrative text comprehension. In other words, children aged 7 years and older may well benefit from illustrations during listening or reading, but they can also do without them.

For educational design and practice, the present results have the following implications. First, for deeper levels of comprehension (i.e., situation model construction), illustrations can be beneficial not only in expository but also in narrative text settings; using them in comprehension problems – especially listening comprehension – can aid students with practicing these tasks. In a more advanced learning step, illustrations may then be removed. Second, illustrations may also help with learning verbal text by rote (i.e., text surface) as long as they do not contribute to extraneous cognitive load, so there should not be too many of them, especially in the reading context where the reading flow could be interrupted, and they should also not be animated. Finally, there are some educational tasks such as re-narrations and content analyses that also require a representation of *what the text says* (i.e., textbase). If illustrations are used for written text in this context, they should be placed before their corresponding text passages so that they do not alter the original textbase representation.

5.6 Limitations and future directions

Of course, the present research has certain methodological drawbacks that should be addressed. First, the short pieces of narrative text used in all three experiments are not typical of the stories to which children in their early school years are exposed. For research on the situation model, this implies that such stories may not allow for discriminating between situation changes that merely reflect details of the plot and situation changes indicating that a text recipient may or may not have grasped the main point of the study. For example, in the story “A Night In The Tent” (Figure 1, Appendix D1), it might be of minor relevance for the further course of action whether the three children take the tent equipment from the basement or attic (original vs. situation change version of sentence 2), but it might be of considerable interest whether they tell each other scary stories or jokes when lying in their sleeping bags (original vs. situation change version of sentence 3). In this spirit, not only the stimulus material used here but also the sentence recognition method as a whole may be of limited usefulness when it comes to discriminating between text recipients who have remembered some details of the state of affairs from those who also – or instead – have understood the main point of the text. Therefore, further studies of text comprehension guided by the tripartite model should use longer and more complex text units and to develop an extended version of Schmalhofer and Glavanov’s (1986) recognition paradigm.

In addition, there should be more empirical attempts to validate the sensitivity measures based on the current sentence recognition method, with more explicit tasks such as text- and inference-based questions like those often used in the multimedia learning literature (e.g., Butcher, 2014) or measures of comprehension-related abilities or skills, which reflects a common practice in the reading comprehension literature (cf. Oakhill, 2020). Regarding the latter, one study (Radvansky & Copeland, 2004) has examined the relationship between the levels of representation assessed via the sentence recognition method and working memory, which has been widely associated with reading comprehension (e.g., Daneman & Carpenter, 1980; Kim, 2015; Seigneuric & Ehrlich, 2005). Radvansky and Copeland (2004) reported that working memory span is positively correlated with sensitivity for the textbase but not the situation model. To my knowledge, there is no other published research that has validated the sentence recognition method against traditional measures of text comprehension and comprehension-related abilities, so future efforts in this area are strongly encouraged.

It could also be objected that illustrations accompanying the limited verbal text are themselves limited to some extent. As already pointed out, the picture-per-word rates were considerably higher than in other studies of narrative text comprehension among children (e.g., O'Keefe & Solman, 1987; Pike et al., 2010) and, importantly, were closer to what would be usual in picture books being read aloud to preschoolers rather than the illustrated storybooks designed for early readers. This may narrow the generalizability of the findings reported here to common reading and listening situations of children in the examined age groups. Moreover, the hand-drawn illustrations used in all three experiments and especially the animations used in Experiment 2 were quite minimalistic and did not contain scene details that could enhance situation model processing on the depictive path of the ITPC; on the other, increased detail could increase cognitive load. In this sense, these findings cannot be generalized easily to representations of other illustrated formats of narrative text or even narrative film. Höffler and Leutner (2007) reported in their meta-analysis that photo-realistic animations (i.e., film material) have a greater positive impact on comprehension than less realistic ones like those included in Experiment 2. Like the static ones, the dynamic illustrations used here were also only related to single sentences and thus disrupted the flow of reading. It can thus be argued that they did not aid participants in keeping track of spatial locations, protagonists, and causal event structures over the course of the story, and it is these features that are considered crucial for situation model construction (Zwaan & Radvansky, 1998). Regarding Experiment 3, it is also possible that static pictures support the perceptual simulation only of static features like shape (Hauf, 2016), whereas it would require dynamic pictures to enhance the perceptual simulation of movement. In summary, future research into mental representations of text-picture units may gain fruitful insights by more systematically varying the properties of illustrations.

A specific limitation of Experiment 3 is that there is no guarantee that the participants actually constructed a situation model of the stories to which they were exposed. The only task they had to perform was judging whether an animated picture presented at the end of the story referred to an object that had appeared in the story. In the 24 critical trials, the target object occurred in the last sentence, so in these cases (the only ones relevant for data analysis), it was sufficient to pay attention to the last sentence rather than the whole story (although there were 6 filler stories where the target object was mentioned before the last sentence). This limitation could be relevant, given that strategies of text reception have an effect on the strength of the situation model (Schmalhofer & Glavanov, 1986). In the event that Experiment 3 participants did not really pay attention to the story plot, the interpretation that they perceptually simulated vertical movement, which is only based on the matching advantage, would not become invalid; on the contrary, this would imply more strongly that perceptual simulation is an automatic and effortless process that occurs during text reception. However, no conclusion can be drawn regarding the functional relationship between perceptual simulation and the situation model when participants do not construct a situation model. Future investigations of the relationship between embodied cognition and the situation model should include measures of both in a single study design.

For three reasons, future investigations on tripartite text representations among children are also recommended to draw inspiration from existing research on reading comprehension. First, reading comprehension research has long been focused on development; this emphasis includes longitudinal study designs and the examination of component skills like working memory (e.g., Oakhill et al., 2003; Oakhill & Cain, 2012; Seigneuric & Ehrlich, 2005). By including these, more insights into the development of text surface, textbase, and situation model representations of text in different presentation formats may be gained. Second, text representations measured via sentence recognition methods like the one used here should be juxtaposed with the explicit measures of comprehension such as memory and inference questions, which are common in the reading comprehension literature. Third, research based on the tripartite model should give more attention to metacomprehension processes such as comprehension monitoring, which has been reported to be an important predictor of text comprehension (e.g., Jaeger & Wiley, 2014; Oakhill et al., 2003; Serra & Dunlosky, 2010; Wiley, 2019).

In summary, the research outlined in this thesis contributes significantly to understanding how children aged 6 to 13 represent and comprehend text presented in illustrated and non-illustrated formats. More aspects taken from the broader sense of *multimedia* can be included in this body of research, with interactive components likely to be of particular interest.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: a cognitive load approach. *Applied Cognitive Psychology*, *21*, 695–700. https://doi.org/10.1002/acp.1343

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*, 1818–1823. https://doi.org/10.1016/j.cub.2006.07.060

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*(2), 85–97. https://doi.org/10.1027//1016-9040.7.2.85

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

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barton, K. (2019). MuMIn: Multi-Model Inference. R package version 1.43.15. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beagles-Roos, J., & Gat, I. (1983). Specific impact of radio and television on children's story comprehension. *Journal of Educational Psychology*, *75*, 128–137.

Beentjes, J.W.J., & van der Voort, T.H.A. (1991). Children's written accounts of televised and printed stories. *Educational Technology Research & Development*, *39*(3), 15–26.

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.

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.

Butcher, K. R. (2006). Learning from text with diagrams: promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*, 182–197. https://doi.org/10.1037/0022-0663.98.1.182

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). New York, NY: Cambridge University Press.

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

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*, 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*(1), 5–26.

Catrambone, R., & Seay, R. F. (2002). Using animation to help students learn computer algorithms. *Human Factors*, *44*(3), 495–511.

Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., deFries, J. C., Wadsworth, S. J., . . . Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: a latent variable analysis. *Journal of Experimental Psychology: General*, *141*(3), 470–488. https://doi.org/10.1037/a0027375

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*, 476–485. https://doi.org/10.1016/j.cognition.2006.02.009

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: use of diagrams in computer-based training environments. *Instructional Science*, *30*, 433–464.

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

De Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

Diakidoy, I.-A.N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, *26*(1), 55–80. https://doi.org/10.1080/02702710590910584

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275–277.

Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, *27*, 153–180. https://doi.org/10.1007/s10648-014-9264-4

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: evidence for scaffolding. *Learning and Instruction*, *28*, 48–63. https://doi.org/10.1016/j.learninstruc.2013.05.002

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. https://doi.org/10.1016/j.jecp.2011.06.009

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences [Computer Software]. *Behavior Research Methods, Instruments & Computers*, *39*, 175–191.

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. https://doi.org/10.1016/j.cognition.2010.05.014

Feierabend, S., Rathgeb, T., & Reutter, T. (2019). KIM-Studie 2018. Kindheit, Internet, Medien. [KIM Study 2018. Childhood, Interet, Media]. Retrieved from https://www.mpfs.de/fileadmin/files/Studien/KIM/2018/KIM-Studie_2018_web.pdf

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: recognition memory for three types of textual information. *Discourse Processes*, *13*, 175–190. https://doi.org/10.1080/01638539009544752

Gambrell, L. B., & Jawitz, P. B. (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*, *28*, 264–276.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.011

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*, 277–300.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 430–445.

Gibbons, J., Anderson, D. R., Smith, R., Field, D. E., & Fischer, C. (1986). Young children's recall and reconstruction of audio and audiovisual narratives. *Child Development*, *57*, 1014–1023.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: pictures help to build mental models. *Journal of Memory and Language*, *31*, 129–151.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379–401. https://doi.org/10.1006/jmla.2000.2714

Greenhoot, A. F., & Semb, P. A. (2008). Do illustrations enhance preschoolers' memories for stories? Age-related change in the picture facilitation effect. *Journal of Experimental Child Psychology*, *99*, 271–287. https://doi.org/10.1016/j.jecp.2007.06.005

Gunter, B., Furnham, A., & Griffiths, S. (2000). Children's memory for news: a comparison of three presentation media. *Media Psychology*, *2*, 93–118. https://doi.org/10.1207/S1532785XMEP0202_1

Guttmann, J., Levin, J. R., & Pressley, M. (1977). Pictures, partial pictures, and young children's oral prose learning. *Journal of Educational Psychology*, *69*, 473–480.

Hald, L. A., van den Hurk, M., & Bekkering, H. (2015). Learning verbs more effectively through meaning congruent action animations. *Learning and Instruction*, *39*, 107–122. https://doi.org/10.1016/j.learninstruc.2015.05.010

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Hauf, J. E. K. (2016). *Experimentelle Untersuchung der Entwicklung der kognitiven Textverarbeitung unter Berücksichtigung unterschiedlicher Darbietungsmodalitäten [Experimental investigation of the development of cognitive text processing with regard to different modalities of text presentations]* (Doctoral Dissertation). Julius-Maximilians Universität, Würzburg.

Hauf, J. E. K., Nieding, G., & Seger, B. T. (2020). The development of dynamic perceptual simulations during sentence comprehension. *Cognitive Processing*, *21*(2), 197–208. https://doi.org/10.1007/s10339-020-00959-7

Hayes, D. S., Kelly, S. B., & Mandel, M. (1986). Media differences in children's story synopses: radio and television contrasted. *Journal of Educational Psychology*, *78*, 341–346.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction*, *17*, 722–738. https://doi.org/10.1016/j.learninstruc.2007.09.013

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*(1), 104–110.

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal*, *32*, 157–163.

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning and Instruction*, *34*, 58–73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Johnston, A. M., Barnes, M. A., & Desrochers, A. (2008). Reading comprehension: Developmental processes, individual differences, and Interventions. *Canadian Psychology*, *49*(2), 125–132. https://doi.org/10.1037/0708-5591.49.2.125

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*, 199–225.

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. https://doi.org/10.1016/j.cognition.2004.06.005

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, *30*, 733–744.

Kaup, B., de la Vega, I., Strozyk, J. V., & Dudschig, C. (2015). The role of sensorimotor processes in meaning composition. In M. H. Fischer & Y. Coello (Eds.), *Conceptual and Interactive Embodiment* (pp. 46–70). London: Routledge.

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior*, *23*, 221–236.

Kim, Y. S. (2015). Language and cognitive predictors of text comprehension: evidence from multivariate analysis. *Child Development*, *86*(1), 128–144. https://doi.org/10.1111/cdev.12293

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension. A Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effect models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227–246). New York, NY: Cambridge University Press.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition, and Neuroscience*, *30*(4), 420–429. https://doi.org/10.1080/23273798.2014.987791

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. https://doi.org/10.1016/j.jphysparis.2008.03.004

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, *81*(2), 240–246.

Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educational Psychologist*, *32*, 1–19.

Mayer, R. E. (2009). *Multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*, 444–452.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*(4), 715–726.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied*, *11*(4), 256–265. https://doi.org/10.1037/1076-898X.11.4.256

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research & Development*, *43*(1), 31–43.

Mayer, R. E. (Ed.) (2014). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43.

Meringoff, L. K. (1980). Influence of the medium on children's story apprehension. *Journal of Educational Psychology*, *72*(2), 240–249.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358–368.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.

Nieding, G. (2006). *Wie verstehen Kinder Texte? Die Entwicklung kognitiver Repräsentationen [How do children understand text? The development of cognitive representation]*. Lengerich: Pabst.

Oakhill, J. (2020). Four decades of research into children's reading comprehension: A personal review. *Discourse Processes*, *57*(5-6), 402–419. https://doi.org/10.1080/0163853X.2020.1740875

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

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*(4), 443–468. https://doi.org/10.1080/01690960344000008

O'Keefe, E. J., & Solman, R. T. (1987). The influence of illustrations on children's comprehension of written stories. *Journal of Reading Behavior*, *19*, 353–377.

Ostarek, M., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, *28*(6), 593–599. https://doi.org/10.1177/0963721419866441

Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*, *182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Paivio, A. (1986). *Mental Representations. A Dual-Coding Approach*. New York: Oxford University Press.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition*, *29*(2), 344–354.

Patson, N. D., George, G., & Warren, T. (2014). The conceptual representation of number. *Quarterly Journal of Experimental Psychology*, *67*(7), 1349–1365. https://doi.org/10.1080/17470218.2013.863372

Perfetti, C., Yang, C. L., & Schmalhofer, F. (2008). Comprehension skill and word-to-text integration processes. *Applied Cognitive Psychology*, *22*, 303–318. https://doi.org/10.1002/acp.1419

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 227–247). Oxford, UK: Blackwell.

Pexman, P. M. (2019). The role of embodiment in conceptual development. *Language, Cognition, and Neuroscience*, *34*(10), 1274–1283. https://doi.org/10.1080/23273798_2017.1303522

Pezdek, K., Lehrer, A., & Simon, S. (1984). The relationship between reading and cognitive processing of television and radio. *Child Development*, *55*(6), 2072–2082.

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, *105*, 243–255. https://doi.org/10.1016/j.jecp.2009.10.006

Pimperton, H., & Nation, K. (2010). Suppressing irrelevant information from working memory: evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language*, *62*, 380–391. https://doi.org/10.1016/j.jml.2010.02.005

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*, 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

R Core Team (2019). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology*, *117*(2), 191–213.

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*(1), 16–20.

Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology*, *94*, 138–144. https://doi.org/10.1037//0022-0663.94.1.138

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. https://doi.org/10.1016/S0364-0213(03)00064-8

Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes*, *41*, 221–250. https://doi.org/10.1207/s15326950dp4103_1

Richter, T., & Zwaan, R. A. (2009). Processing of color words activates color representations. *Cognition*, *111*, 383–389. https://doi.org/10.1016/j.cognition.2009.02.011

Rideout, V., & Robb, M. B. (2019). *The common sense census: media use by tweens and teens*. San Francisco: Common Sense Media.

Rieber, L. P. (1990). Using computer animated graphics in science instruction with children. *Journal of Educational Psychology*, *82*(1), 135–140.

Ruggeri, A., & Katsikopoulos, K. V. (2013). Make your own kinds of cues: when children make more accurate inferences than adults. *Journal of Experimental Child Psychology*, *115*, 517–535. https://doi.org/10.1016/j.jecp.2012.11.007

Salomon, G. (1984). Television is "easy" and print is "tough": the differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*(4), 647–658.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: verbatim, propositional and situational representations. *Journal of Memory and Language*, *25*, 279–294.

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: a comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*, 306–312.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 72–103). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Seigneuric, A., & Ehrlich, M. F. (2005). Contribution of working memory capacity to children's reading comprehension: a longitudinal investigation. *Reading and Writing*, *18*, 617–656. https://doi.org/10.1007/s11145-005-2038-0

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory*, *18*, 698–711. https://doi.org/10.1080/09658211.2010.506441

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2010). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Cengage.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Speed, L. J., & Vigliocco, G. (2014). Eye movements reveal the dynamic simulation of speed in language. *Cognitive Science*, *38*, 367–382. https://doi.org/10.1111/cogs.12096

Springer, K. (2001). Perceptual boundedness and perceptual support in conceptual development. *Psychological Review*, *108*(4), 691–708. https://doi.org/10.1037//0033-295X.108.4.691

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*, 137–149.

Sticht, T. G., Beck, L. J., Hauke, R. N., Kleiman, G. M., & James, J. H. (1974). *Auding and Reading: A Developmental Model*. Alexandria, VA: Human Resources Research Organization.

Strozyk, J. V., Dudschig, C., & Kaup, B. (2019). Do I need to have my hands free to understand hand-related language? Investigating the functional relevance of experiential simulations. *Psychological Research*, *83*, 406–418. https://doi.org/10.1007/s00426-017-0900-8

Taylor, L. J., & Zwaan, R. A. (2008). Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology*, *61*, 896–904. https://doi.org/10.1080/17470210701625519

Taylor, L. J., & Zwaan, R. A. (2009). Action and cognition: the case of language. *Language and Cognition*, *1*(1), 45–58. https://doi.org/10.1515/LANGCOG.2009.003

Therriault, D. J., Yaxley, R. H., & Zwaan, R. A. (2009). The role of color diagnosticity in object recognition and representation. *Cognitive Processing*, *10*, 335–342. https://doi.org/10.1007/s10339-009-0260-4

Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., & Rapp, D. N. (2009). Simple but complex: components of the simple view of reading across grade levels. *Journal of Research in Reading*, *32*(4), 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x

Unsöld, I. H., & Nieding, G. (2009). Die Bildung prädiktiver Inferenzen von Kindern und Erwachsenen bei der kognitiven Verarbeitung audiovisueller und auditiver Texte. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *41*(2), 87–95. https://doi.org/10.1026/0049-8637.41.2.87

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

Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process, and development. *Applied Cognitive Psychology*, *22*, 293–301. https://doi.org/10.1002/acp.1417

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*, 407–423. https://doi.org/10.1002/acp.1414

Walma van der Molen, J. H., & van der Voort, T. H. A. (2000). The impact of television, print, and audio on children's recall of the news. *Human Communication Research*, *26*(1), 3–26.

Wannagat, W., Waizenegger, G., Hauf, J., & Nieding, G. (2018). Mental representations of the text surface, the text base, and the situation model in auditory and audiovisual texts in 7-, 9-, and 11-year-olds. *Discourse Processes*, *55*(3), 290–304. https://doi.org/10.1080/0163853X.2016.1237246

Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing*, *18*, 491–504. https://doi.org/10.1007/s10339-017-0820-y

Wiley, J. (2019). Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Applied Cognitive Psychology*, *33*, 9–19. https://doi.org/10.1002/acp.3495

Winter, B., & Bergen, B. (2012). Language comprehenders represent object distance both visually and auditorily. *Language and Cognition*, *4*(1), 1–16. https://doi.org/10.1515/langcog-2012-0001

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*, 229–236. https://doi.org/10.1016/j.cognition.2006.09.003

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88. https://doi.org/10.1080/01638539909545070

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*, 229–234. https://doi.org/10.1016/j.tics.2014.02.008

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. https://doi.org/10.1016/j.cogsci.2004.03.004

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PLoS ONE*, *7*(12), e51382. https://doi.org/10.1371/journal.pone.0051382

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

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

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 & Language*, *112*, 143–149. https://doi.org/10.1016/j.bandl.2008.11.004

Appendix

Appendix A: Representations of written text (Experiment 1)

This is the accepted manuscript version of the following research article: Seger, B. T., Wannagat, W. & Nieding, G. (2021). Children's surface, textbase, and situation model representations of written and illustrated written narrative text. *Reading and Writing, 34*, 1415-1440. doi:10.1007/s11145-020-10118-1

Abstract

According to the tripartite model of text representation (van Dijk & Kintsch, 1983), readers form representations of the text surface and textbase, and construct a situation model. In this study, an experiment was conducted to investigate whether these levels of representation would be affected by illustrations added to narrative text and whether the order of text and illustrations would make a difference. Students aged between 7 and 13 years ($N = 146$) read 12 narrative texts, four of them with illustrations presented before their corresponding sentences, four with illustrations presented after, and four without any illustration. A sentence recognition task was used to assess the accuracy for text surface, textbase, and situation model. For the text surface and situation model, neither the presence of illustrations nor the order of text and illustrations influenced accuracy. However, the textbase was negatively affected by illustrations when they followed their corresponding sentences. We suggest that after situation model construction, illustrations can initiate model inspection (Schnotz, 2014), a process that can make substantial changes to the textbase representation.

Introduction

Generations of children have been exposed to illustrated storybooks when tales were read aloud by the children's caregivers. To date, much research has been conducted showing a functional link between reading from storybooks and children's language comprehension and literacy development (e.g., Duursma, Augustyn, & Zuckerman, 2008; Isbell, Sobol, Lindauer, & Lowrance, 2004; Klein & Kogan, 2013). Illustrations in storybooks seem to play a crucial role during the activity of reading aloud, and young children are supposed to heavily rely on the information conveyed by the illustrations during story retelling (Isbell et al., 2004). Books and novels for older, literate children also often include illustrations, albeit to a lesser extent than storybooks for younger children. Certainly, these have an ornamental function, but the question arises regarding whether and how illustrations may also contribute to the understanding of narrative content during silent reading.

Several experiments reveal that children recall narrative text better and generate more appropriate inferences when the verbal text is accompanied by appropriate illustrations (e.g., Beagles-Roos & Gat, 1983; Beentjes & van der Voort, 1991; Gambrell & Jawitz, 1993; Gibbons, Anderson, Smith, Field, & Fischer, 1986; Greenhoot & Semb, 2008; Guttmann, Levin, & Pressley, 1977; Hayes, Kelly, & Mandel, 1986; O'Keefe & Solman, 1987; Pike, Barnes, & Barron, 2010; Ricci & Beal, 2002; Salomon & Leigh, 1984; for a review see Pressley, 1977). In some studies, differences between verbal-and-visual and verbal-only media are more pronounced in younger than in older children (Gibbons et al., 1986; Guttmann et al., 1977; Pike et al., 2010). Our research goal is to specify how illustrations are related to both superficial and deeper comprehension levels of written narrative text. To this end, we refer to a theoretical account that provides three levels of text representation, as well as to models of multimedia learning that use this theory to explain the comprehension of both verbal text and text-picture units.

We use *text* as an umbrella term including every presentation modality (written, auditory, and audiovisual) and genre (narrative and expository) of text this work refers to[1]; if applicable, *text* refers to the ensemble of words (verbal text) and pictures. We define *stories* as coherent units of verbal narrative text of any length. The term *picture* encompasses any nonverbal, visual text

[1] At this point, it seems to be worth clarifying that none of the research referred to in this study systematically uses text that can be assigned to the linguistic concept of *spoken language*, which has, among other features, a different communicative function compared with written text. Auditory verbal text is usually conceived as a spoken version of written text.

elements that can have different functions in connection with verbal text (e.g., schematic representation, metaphor, additional information, illustration). The concept of *illustration* is exclusively used in the context of narrative text and refers to pictures that repeat what happens in the story, without adding information that is relevant to understand the situation that the story refers to. Nonetheless, illustrations may contain details of the scene that are not explicitly mentioned in the story.

Text surface, textbase, and situation model

The tripartite theory of text comprehension (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) claims that text recipients form three different mental representations of verbal text: text surface, textbase, and situation model. The *text surface* refers to the exact wording, whereas the *textbase* covers the semantic content that can be seen as a network of propositions (Kintsch, 1988, 1998). Propositions are defined as the smallest meaning units to which a truth value can be assigned, and they are usually outlined using predicate-argument structures (e.g., Engelkamp, 1980). A sentence such as *Jane is watering the flowers in the garden* may be expressed as WATER(agent: Jane; object: flowers; location: garden). If the sentence is framed in the passive voice, such as *The flowers in the garden are being watered by Jane*, the textbase remains identical, whereas the text surface is different.

Whereas the text surface and textbase are closely tied to the verbal text, the *situation model* is a coherent representation of the situation referred to in the text and is constructed by drawing inferences. For example, if one reads the sentence mentioned above, one may infer that Jane feels responsible for the flowers or that it has not been raining for several days. Researchers who follow an embodied cognition account (e.g., Barsalou, 1999) further claim that situation models can be described as analogous, multidimensional, and modality-specific *simulations* of real-world events. These simulations are supposed to be largely based on the recipient's perceptual and motor experience (Glenberg & Robertson, 2000; Stanfield & Zwaan, 2001; Taylor & Zwaan, 2009; Zwaan, 1999, 2014; Zwaan, Stanfield, & Yaxley, 2002). While reading the sentence *Jane is watering the flowers in the garden*, one may easily imagine seeing the flowers' colors, smelling their fragrance, or hearing the water pouring out of the watering can.

A sentence recognition method has been developed to establish all three representations of verbal text at once (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986). The above-cited researchers found that surface, textbase, and situation model representations simultaneously occurred among adults. The participants were able to discriminate between an original sentence and a *paraphrase*, where the exact wording, but not the propositional structure, had changed.

They were observed to discriminate better between paraphrases and *meaning changes*, where the propositional structure was also altered while remaining true to the situation (e.g., Jane is watering the flowers *outside*). However, discrimination was best when a *situation change* was presented in a sentence that should also be incompatible with the recipient's situation model (e.g., Jane is watering the flowers *on the balcony*). Author (2006) replicated this pattern of results in a sample of 5–11 year-old children, so there is evidence that the tripartite model appropriately describes text comprehension in childhood.

Based on this, we examined whether illustrations would make a difference in elementary school students' comprehension of auditory narrative text (Seger, Wannagat, & Nieding, 2019; Wannagat, Waizenegger, Hauf, & Nieding, 2018; Wannagat, Waizenegger, & Nieding, 2017). Wannagat et al. (2018) asked their 7, 9 and 11 year-old participants to listen to stories that comprised six sentences each before they completed a sentence recognition task where they were presented with original sentences, paraphrases, meaning changes, and situation changes. In one experimental condition, the participants received these stories in an auditory-only version; in the other condition, every sentence was accompanied with a static illustration. Similarly, Seger et al. (2019) scrutinized text surface, textbase, and situation model representations of auditory and audiovisual stories in the same group and with roughly the same stimulus material, with the exception that they added a third experimental condition where animated instead of static illustrations were used.

In both studies, the situation model was significantly improved when illustrations were present rather than absent; likewise, text surface representations seemed to benefit from illustrations. One study (Wannagat et al., 2018) revealed an opposite pattern of results at the textbase level, indicating that semantic representations of text are less accurate when the text is illustrated; this was not replicated in Seger et al.'s (2019) research. The latter also suggested that dynamic illustrations produce similar results as static ones when accompanying auditory narrative text. To our knowledge, the effect of illustrations on the comprehension of written narrative text has not yet been investigated with reference to the tripartite model.

Theories of text-and-picture processing

Based on a large body of research on expository text comprehension, Mayer (1997, 2009) formulated a *multimedia principle*, which states that people learn better from verbal text with pictures than from verbal text alone. In this research tradition, all media that present words and pictures are referred to as multimedia, and multimedia learning is defined as building mental representations from words and pictures. In her review, Butcher (2014) showed that the multimedia principle is applicable to a variety of learning forms, including both superficial and deep levels of learning, and to a variety of media types.

Comparisons between expository text with and without pictures were shown to favor the multimedia principle, especially with regard to deep level learning. Glenberg and Langston (1992), for example, found that mental models based on written expository text improved when pictures were provided. Similar effects were obtained in a training study with hypermedia (Cuevas, Fiore, & Oser, 2002); the participants performed better in an integrative knowledge task when pictures were included in the hypermedia, but not in a declarative knowledge task. The pictures in both studies were schematic diagrams that organized the information provided by the text without containing additional information. Butcher (2006) additionally varied between simplified (conceptually true) and complex (physically true) diagrams. Results suggested that pictures improve the mental modeling of expository text and that simple diagrams do more so than complex ones. The latter effect is explained by the notion that pictures have a beneficial effect on mental modeling when they can highlight crucial information from the text (i.e., provide a visual summary). In addition, participants in the simple diagram condition outperformed those in other conditions regarding memory of details.

The integrated model of text and picture comprehension (ITPC; Schnotz, 2014; Schnotz & Bannert, 2003) uses van Dijk and Kintsch's tripartite model to explain the multimedia principle. The ITPC assumes that processing text-picture units involves two channels: (1) a descriptive one proceeding from verbal text and (2) a depictive one proceeding from pictures. Accordingly, text surface representation arises from sub-semantic processing, and textbase representation emerges from semantic processing on the descriptive path. In contrast, the situation model is a depictive representation of the text and can be acquired in two ways: (1) *situation model construction* (e.g., Zwaan & Radvansky, 1998) based on semantic information gathered from descriptive processing (textbase) and one's own knowledge of the world and/or (2) *analog structure mapping* (Gentner, 1989) based on the picture surface representation, which is directly gathered on the depictive path. If the picture reproduces central features of its corresponding verbal text (this includes *illustrations*, according to our working definition), analog structure mapping can be used to match a constructed situation model with the picture surface representation because they are both depictive representations and thus share structural properties.

Analogue structure mapping can explain why situation models improve when audiovisual text rather than auditory-only text has been presented (e.g., Author, 2019). It can also be argued that analogue structure mapping reduces the need for semantic processing, which may result in lower textbase representations in the presence of pictures (Author, 2018). Moreover, Schnotz and Bannert (2003) proposed that text recipients can apply *model inspection* processes after they have constructed a situation model. In doing so, they obtain new information from the situation model and encode this information in a propositional format. Such new information can have its origin in an illustration of verbal text; as a consequence, pictorial information may be encoded into propositions via model inspection, meaning that pictures may interfere with textbase representations.

Impact of pictures on the comprehension of written text

In the domain of narrative text, there is empirical evidence that illustrations support the comprehension of written stories. Gambrell and Jawitz (1993) examined the recall of four-page stories with versus without illustrations in a sample of 10 year-old children. Participants who read illustrated stories outperformed those reading verbal-only stories in both free and probed recall measures. Similar results were obtained by O'Keefe and Solman (1987) with stories comprising about 470 words (approximately one A4 page). Recall accuracy was higher when the presentation of the story and illustrations was sequential (experiments 1 and 2) or simultaneous (experiment 3). Whereas both of these studies used relatively long text units with low picture-per-word rates (less than 1:100), the stories used by Pike et al. (2010) contained five sentences (60—70 words) with one illustration per story, resulting in a somewhat higher picture-per-word rate. Pike et al. found that readers aged between 7 and 10 years drew more correct inferences when illustrations were present rather than absent.

Whereas the multimedia principle is insensitive to the modality in which verbal text is presented (auditory vs. written), the *modality principle* (Low & Sweller, 2014; Moreno & Mayer, 1999; Mousavi, Low, & Sweller, 1995) claims that multimedia learning benefits more from a text that employs two sensory channels (auditory-visual) rather than one (visual-visual). A somewhat intuitive explanation of the modality principle would be that audiovisual text can be simultaneously encoded on two sensory channels, whereas the early visual processing of written text and pictures has to be successive, thereby creating a bottleneck. However, there has been a debate about where exactly this bottleneck occurs. The *split-attention effect*, for instance, claims that this bottleneck concerns attentional focus and can thus be overcome by spatially integrating written text and pictures (e.g., using diagram labeling; Ayres & Sweller, 2014). Alternatively, Rummer and colleagues (Rummer, Schweppe, Fürstenberg, Scheiter, & Zindler, 2011; Rummer, Schweppe, Fürstenberg, Seufert, & Brünken, 2010) introduced a *sensory register* hypothesis (see also Penney, 1989) claiming that a pre-attentive integration of verbal text and picture would be easier with auditory than with written text (for a critical discussion, see Reinwein, 2012). Ascribing the visual-visual bottleneck to early sensory processing would also be in line with the ITPC (Schnotz, 2014), according to which the sub-semantic but not the semantic processing stage can be affected by such a bottleneck because the latter provides that phonetic decoding of written text has already taken place.

Nonetheless, it can be helpful to examine the possible effects of the processing order while investigating the comprehension of written text-picture units. Experimentally varying the presentation order of text and pictures seems to be a plausible way to do so. In the field of expository text comprehension, Eitel and Scheiter (2015) conducted a systematic review of the studies that performed this variation. In the studies they reported, the number of findings indicating that comprehension was better when the text was shown before the picture (e.g., Canham & Hegarty, 2010) was almost equal to the number of findings that revealed the opposite pattern (e.g., Baggett, 1984; Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013). As far as we know, in the domain of narrative text, only one attempt has been made to directly assess whether the order of text and pictures affects comprehension. A combined analysis of experiments 1 and 2 in O'Keefe and Solman’s (1987) study indicated that illustrations presented before or after their corresponding story improved recall compared with verbal-only text, but there was no difference regarding the order of verbal text and illustrations.

This study

The aim of our study was to understand how illustrations affect children’s comprehension of written stories and to examine whether the processing order of verbal text and illustrations would make a difference. In particular, we investigated how each level of representation according to the tripartite model (the text surface, textbase, and situation model) would be affected (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). To obtain separate measures for each level, we employed a sentence recognition task similar to the one introduced by Schmalhofer and Glavanov (1986) and used in several further experiments (Author, 2006, 2017, 2018, 2019; Fletcher & Chrysler, 1990). The stories in our study reflected possible daily-life situations of school children in Western countries. We experimentally varied three story versions: written stories without illustrations *(text-only)*, written stories with illustrations presented before it *(picture-text, PT)*, and written stories with illustrations presented after it *(text-picture, TP)*. Another purpose of our study was to examine if beginning readers at age 7 would differ from more advanced readers up to age 13 in their comprehension of written and illustrated written narrative text. Finally, we studied the effects of illustrations and the text-illustration order on reading times.

We anticipated that the *situation model* would benefit from illustrations in general, consistent with multimedia learning theories (Mayer, 2009; Schnotz, 2014) as well as earlier results from both auditory (Beagles-Roos & Gat, 1983; Gunter, Furnham, & Griffiths, 2000; Hayes et al., 1986; Seger et al., 2019; Wannagat et al., 2018) and written narrative text (Gambrell & Jawitz, 1993; Pike et al., 2010). We also assumed that situation model representations would be more accurate in the PT than in the TP condition. This would be in line with the ITPC (Schnotz & Bannert, 2003) claiming that an appropriate situation model could be directly obtained via analog structure mapping, which serves as a scaffold for the subsequent, more complex process of situation model construction based on verbal text. Thus, Hypothesis 1 predicted the accuracy to be in the order of PT > TP > text-only for the situation model.

Regarding the *textbase*, we expected illustrations to have a negative effect. We derived this assumption from the ITPC. If the situation model could be directly obtained from a picture surface representation, semantic processing might become less relevant to this objective and might therefore be neglected. This effect was found in one of our earlier studies with auditory stories (Wannagat et al., 2018), but not in others (Seger et al., 2019; Wannagat et al., 2017). In addition, new information obtained from an illustration could alter textbase representations via model inspection (Schnotz & Bannert, 2003). As model inspection is presumed to take place after model construction, we thought that this effect would be more likely when the illustration would be presented after the written text rather than before it. Therefore, Hypothesis 2 predicted that accuracy would be lower when illustrations were present rather than absent and that accuracy in TP would be lower than that in PT (i.e., text-only > PT > TP for the textbase).

For the *text surface*, we hypothesized that illustrations would have a positive effect, consistent with our earlier results with auditory versus audiovisual text (Seger et al., 2019; Wannagat et al., 2018). However, we made no assumption regarding the order of text and illustrations (Hypothesis 3: TP = PT > text-only). Hypothesis 4 claimed that illustrations would facilitate subsequent reading, which would be reflected in lower reading times when illustrations were present in general and when they were presented before the written text in particular (PT < TP < text-only for reading time).

Method

Participants

We determined an ideal sample size of $N = 144$ that would enable an optimal balance across participants and conditions (see below for more details). A power analysis conducted via G*Power (Version 3.1.9.2, Faul, Erdfelder, Lang, & Buchner, 2007) indicated that with this sample size, a true effect size of $\eta^2 = .020$ would be detected with a likelihood of more than 90% (i.e., $\beta < .10$). This effect size is remarkably below the effect sizes associated with the significant results obtained in earlier sentence recognition studies (ranging between $\eta^2 = .040$ and $\eta^2 = .092$; Wannagat et al., 2017, 2018; Seger et al., 2019).

In total, 146 students aged between 7.75 and 13 years (mean age = 10.42, $SD = 1.25$, median = 10.58) participated in our study, with females comprising a slight majority (53%). The participants were recruited from several elementary schools and a comprehensive secondary school in Germany. All participants spoke German at the native-speaker level. The students only participated after their parents had signed a consent form.

Sentence recognition task

We used a three-level sentence recognition task based on the method introduced by Schmalhofer and Glavanov (1986). Our task is an adapted version of the one used in earlier studies with children (Wannagat et al., 2017, 2018; Seger et al., 2019). The participants read stories composed of six sentences each. After a block of four stories, they read single sentences and were required to decide whether these were part of the story. The sentences were either presented in their original wording, requiring a positive answer, or were modified in one of three ways: as a *paraphrase*, where the wording (i.e., text surface) was changed without changing the meaning at the sentence level (e.g., by replacing one or more expressions with synonyms); as a *meaning change*, where the meaning at the sentence level (i.e., textbase) was altered but remained true to the story plot; or as a *situation change*, where the meaning of a sentence was modified in a way that was incompatible with the plot (i.e., meant to contradict the reader's situation model).

Twelve stories were included in the task (see Figure 1 for an example); these were related to everyday events that might occur in a child's life in Western cultures to ensure that domain-specific knowledge or expertise would not be necessary. Text coherence was ensured locally by employing theme-rheme structures (e.g., pronouns that unambiguously refer to a character or object occurring in the previous sentence) and globally by providing an appropriate title in advance (Bransford & Johnson, 1972), which appeared in capital letters[2]. A vast majority (91.7%) of the sentences described one or more characters' actions; some sentences (31.9%) referred to a character's emotional state. For each original sentence, three distractors were created that met the criterion of paraphrase, meaning change, and situation change, respectively (for an example see Table 1). The sentence length varied between 10 and 22 words (mean = 15.23, *SD* = 2.47, median = 15) with negligible differences between sentence types. In the two illustrated conditions, one static illustration preceded or followed every sentence. Most of them depicted at least one character (90.3%) and/or action or emotional state (87.5%) the corresponding sentence referred to. We ensured that the illustrations did not include any detail that might be incompatible with the distractor sentences, especially the situation change versions.

During the task, six probe sentences were presented in scrambled order: three as original sentences, one as a paraphrase, one as a meaning change, and one as a situation change. The probe sentences were balanced as much as possible in two ways. First, we ensured that for each of the 72 sentences, every sentence type appeared equally often among all participants and in each condition (i.e., each sentence appeared equally often in the paraphrase, meaning change, and situation change versions, respectively; each sentence appeared in the original version as frequently as in all changed versions together). Second, we ensured that the position of each sentence in the task was equally distributed (e.g., the first sentence of a story was equally often the first, third, or last sentence in its related task).

The verbal text was presented in black Arial font in the top third of a white 800 * 600 pixel field; the font size was 20 points for sentences and 26 points for titles. The illustrations were hand-drawn and colored (see Figure 1), with a uniform size of 800 * 600 pixels. The experiment was implemented in the DMDX® software, Version 5 (Forster & Forster, 2016) on a laptop computer, with a resolution of 1280 * 720 and frame rate of 60 Hz.

[2] We employed theme-rheme structures and titles to the best of our knowledge and belief, but we did not empirically test their adequacy. We also ensured that the titles did not interfere with any of the distractor sentences.

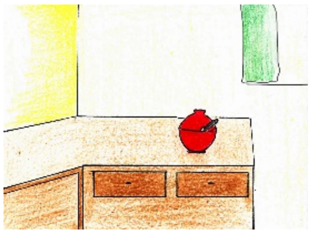
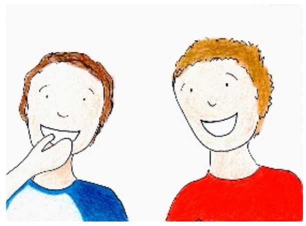
No.	*Original sentence in German*	English translation	Illustration
1	*Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade ihre Lieblingsserie anschauen.*	Mother calls out for Max and Sascha, but the two of them just wanted to watch their favorite serial.	
2	*Die Oma kommt zum Mittagessen vorbei und Max und Sascha sollen deshalb den Tisch decken.*	Grandma comes around for lunch, therefore, Max and Sascha are supposed to set the table.	
3	*Als sie widerwillig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.*	As they unwillingly fetch up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard.	
4	*Grinsend zwinkert er Sascha zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein.*	Smirkingly, he twinkles to Sascha and then pours the sugar from the red bowl into the salt shaker.	
5	*Bald sitzen alle am Tisch und Max und Sascha lassen den Salzstreuer nicht aus den Augen.*	Pretty soon, everyone sits at table and Max and Sascha keep their eyes glued to the salt shaker.	
6	*Als der Vater dann endlich nach dem Salzstreuer greift, können die beiden ihr Lachen nicht mehr unterdrücken.*	As Father finally reaches for the salt shaker, the two of them can no longer keep under their laughter.	

Figure 1. Sample story entitled *Beim Essen* (At Lunch) and its illustrations.

Table 1
Original Sentences, Paraphrases, Meaning Changes, and Situation Changes of the Third Sentence from the Story Beim Essen *(At Lunch)*

Sentence type	German version used in the study	English translation
Original	Als sie widerwillig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	As they unwillingly fetch up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard.
Paraphrase	Als sie wider*strebend* Teller und Gläser aus der Küche herbei*bringen*, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	As they *reluctantly bring* up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard.
Meaning Change	Als sie widerwillig *Geschirr* aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	As they unwillingly fetch up *dishes* from the kitchen, Max suddenly sees a red sugar bowl on the sideboard.
Situation Change	Als sie *eifrig* Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	As they *eagerly* fetch up plates and glasses from the kitchen, Max suddenly sees a red sugar bowl on the sideboard.

Design and procedure

Three experimental conditions were varied within participants: one text-only, one with illustrations presented before their corresponding sentences (PT) and one with illustrations presented after (TP). The participants read the 12 stories in three blocks of four stories each, where each block represented one condition. All possible orders of experimental conditions were permutated and randomly assigned to the participants; however, we tried to balance them in terms of age, gender, and day time (class hours) as far as possible.

For the experimental task, the students were instructed to read the stories and remember them as accurately as possible. Concerning the sentence recognition task, they were instructed to expect a test where they would be presented with sentences in arbitrary order and would have to decide whether these sentences appeared in one of the stories. For “yes”, they pressed the “3” key on the numeric keyboard which was stickered with a happy emoticon, for “no”, they pressed the “1” key stickered with a sad emoticon. They completed a practice trial comprising three sentences and three respective probes in the following order: situation change, original, and paraphrase. We provided no feedback at any time. However, after the practice trial, we asked the participants whether they had understood how to perform the task, and we repeated the instructions if the response pattern in the practice trial raised the issue that the participants might not have correctly understood them (e.g., if they considered the order of sentences during the task). During reading, the participants always proceeded by pressing the “Enter” key (stickered with a book symbol) for the next sentence or picture to appear; thus, reading and picture-viewing were self-paced with no time limit. The reading and picture-viewing times were automatically measured by the experimental software. The task phase also had no time limit except for the titles, each of which was shown for three seconds and served as a reminder for the respective story. No pictures were shown during the task phase. When a reading block was completed (after four stories), a short instruction announcing the task phase appeared in red. After the task, a short instruction appearing in green announced the next or the last block or the end of the experiment. The entire experiment usually required 25—40 minutes.

Results

Preliminary analyses and sensitivity calculation

We calculated the acceptance rates (i.e., the relative frequencies of "yes" responses) for originals, paraphrases, meaning changes, and situation changes to determine whether the tripartite model would be appropriate to describe text comprehension in our study. We considered this to be the case if the acceptance rates were the highest for originals and decreased with increasing change intensity. The mean acceptance rate was 0.862 (SD = 0.103) for originals, 0.744 (SD = 0.167) for paraphrases, 0.602 (SD = 0.179) for meaning changes, and 0.287 (SD = 0.178) for situation changes (see Table 2). A repeated-measure analysis of variance (ANOVA) exhibited a significant effect of the sentence type, $F(3, 143) = 321.91$, $p < .001$, $\eta^2 = .871$. Contrast analyses showed significant differences between originals and paraphrases, $F(1, 145) = 72.39$, $p < .001$, $\eta^2 = .333$, paraphrases and meaning changes, $F(1, 145) = 70.47$, $p < .001$, $\eta^2 = .327$, as well as meaning changes and situation changes, $F(1, 145) = 358.982$, $p < .001$, $\eta^2 = .712$. Thus, we assumed that the tripartite model was applicable to the sentence recognition task in our sample.

As a reliability measure for our sentence recognition task, we determined Cronbach's alpha for each acceptance rate, which reflects its internal consistency across stories. The value was in an acceptable range for original sentences ($\alpha = .708$) but not for paraphrases ($\alpha = .476$), meaning changes ($\alpha = .423$), and situation changes ($\alpha = .512$).

For each level of representation, sensitivities based on the signal detection theory (Stanislaw & Todorov, 1999) were computed. We deemed this necessary because the acceptance rates of a certain change type do not unambiguously refer to the respective level of representation. For instance, accepting a situation change as being part of the story indicates that the reader had not constructed an appropriate situation model; however, rejecting a situation change can also indicate that a reader merely had a correct representation of the text surface or textbase, as situation changes necessarily imply meaning changes and meaning changes necessarily imply paraphrases. Moreover, sensitivity measures have the advantage that they are independent of the recipient's response bias (criterion; see Stanislaw & Todorov, 1999).

We used the A' sensitivity measure that is nonparametric (i.e., does not require normally distributed values; Donaldson, 1992) and ranges from 0 to 1, with 0.5 representing the chance level. For text surface A', "yes" responses to originals were categorized as hits and "yes" responses to paraphrases were categorized as false alarms (i.e., false positives). For textbase A', "yes" responses to originals and paraphrases were considered hits and "yes" responses to meaning changes were considered false alarms. Finally, for the situation model, "yes" responses to originals, paraphrases, and meaning changes were regarded as hits and "yes" responses to situation changes were regarded as false alarms. In general, we assigned the acceptance of a specific change type to false alarms, indicating that the subject had no adequate text representation at the corresponding level; moreover, we designated the combined acceptance rates at the more superficial levels as hits (see also Seger et al., 2019). For the formulas, see Figure 2; for the descriptive statistics, see Table 2. Please note that *A'* cannot be expressed as a real number if the hit rate is zero or the false alarm rate is one. If such a case occurred in at least one experimental condition, the participants were excluded from hypothesis testing at the corresponding text comprehension level. Therefore, we specified the number of participants included in our analyses of text surface, textbase, and situation model sensitivities.

The sample parameters for reading and picture-viewing times are described in Table 2. Not surprisingly, reading times were negatively correlated with age ($r = -.322$, $p < .001$). Sensitivity measures and acceptance rates did not correlate with reading or picture viewing times ($|r| \leq .146$, $p \geq .079$), indicating that there is no speed-accuracy tradeoff in our data. Sensitivity measures and acceptance rates were also unrelated to age.

General
$$A' = 0.5 + \frac{(H - F)(1 + H - F)}{4H(1 - F)}$$

Surface
$$A' = 0.5 + \frac{(Y_O - Y_P)(1 + Y_O - Y_P)}{4Y_O(1 - Y_P)}$$

Textbase
$$A' = 0.5 + \frac{(Y_{O,P} - Y_M)(1 + Y_{O,P} - Y_M)}{4Y_{O,P}(1 - Y_M)}$$

Situation model
$$A' = 0.5 + \frac{(Y_{O,P,M} - Y_S)(1 + Y_{O,P,M} - Y_S)}{4Y_{O,P,M}(1 - Y_S)}$$

Figure 2. Formulas for the nonparametric Signal Detection sensitivity measures (*A'*s) used in our study. H = hit rate, F = false alarm rate, Y = acceptance rate, O = original sentence, P = paraphrase, M = meaning change, S = situation change

Levels of representation

Because sensitivity measures showed correlation with age, we excluded it in the analyses for levels of representation. We also did not calculate a multivariate ANOVA that would allow for direct comparisons between levels of representation owing to the statistical interdependencies between the sensitivity measures. Thus, repeated-measure ANOVAs with the text format as predictor were separately performed for text surface, textbase, and situation model sensitivities.

If *A'* was not calculable for at least one experimental condition (see above), the participants were excluded from the analyses at the corresponding level of representation. This was the case for 37 participants (25.3%) for the surface, seven participants (4.8%) for the textbase, and one participant (0.7%) for the situation model analyses.

For the situation model, the effect of the text format was not significant, $F(2, 143) = 0.272$, $p = .763$, leading to the refutation of our assumption that illustrations would enhance situation model representations of written narrative text (Hypothesis 1). However, a significant effect emerged at the textbase level, $F(2, 137) = 7.958$, $p = .001$, $\eta^2 = .104$. Planned contrasts revealed significantly higher accuracies in PT than in TP, $F(1, 138) = 15.605$, $p < .001$, $\eta^2 = .102$, whereas there was no significant difference between both illustrated conditions and the text-only condition, $F(1, 138) = 0.624$, $p = .431$. This partly corroborates Hypothesis 2; accuracy was significantly higher when the picture was presented before the text rather than after, but there was no general advantage of the text-only condition over both illustrated conditions. Text surface *A'* was not affected by the text format, $F(2, 107) = 1.084$, $p = .342$; therefore, Hypothesis 3 was rejected. The descriptive statistics for sensitivities in dependence on experimental conditions are summarized in Table 3.

Reading and picture viewing times

As reading time was significantly related to age, we ran an analysis of covariance (ANCOVA) to determine a possible interaction between experimental conditions and age. This interaction was not significant, $F(2, 143) = 1.371$, $p = .257$; therefore, we decided to perform an ANOVA instead. The effect of the text format on reading times was significant, $F(2, 144) = 5.562$, $p = .005$, $\eta^2 = .072$. Planned contrasts indicated shorter reading times in the illustrated conditions than in the text-only condition, $F(1, 145) = 9.577$, $p = .002$, $\eta^2 = .062$, whereas the contrast between the PT and TP conditions did not reach significance, $F(1, 145) = 0.332$, $p = .565$. These findings confirmed Hypothesis 4 insofar as reading times differed between the text-only and both illustrated conditions but not between PT and TP. Unexpectedly, illustrations were viewed longer when they followed rather than preceded their corresponding sentences, $t(145) = 2.125$, $p = .035$. For an overview of reading and picture-viewing times in dependence on text format, see Table 3.

Analyses for carryover effects

Although the order of experimental conditions was balanced across participants, we were interested in the carryover effects that may have occurred between them. To this end, we re-ran our analyses of text surface, textbase, and situation model *A'*s with an additional between-participant factor indicating which text condition was completed first (text-only vs. PT vs. TP). This factor yielded a significant main effect for the textbase, $F(2, 136) = 3.155$, $p = .046$, $\eta^2 = .044$; however, Bonferroni-adjusted post-hoc comparisons did not reveal significant group differences for this factor. More interestingly, a significant interaction was observed between this factor and the experimental factor for the textbase, $F(4, 272) = 3.257$, $p = .012$, $\eta^2 = .046$. Bonferroni-adjusted post-hoc comparisons indicated significantly lower textbase *A'*s in the TP condition than in the text-only (mean difference = 0.140, $p = .010$) and PT conditions (mean difference = 0.204, $p < .001$) in the group of participants who started with the text-only condition. Participants who started with the PT condition yielded higher textbase *A'*s in the PT condition than in the TP condition (mean difference = 0.122, $p = .016$), whereas there were no significant differences between the conditions related to the group starting with the TP condition.

For the situation model, this interaction was also significant, $F(4, 284) = 6.373$, $p < .001$, $\eta^2 = .082$. We ran Bonferroni-adjusted post-hoc comparisons suggesting higher performance in the text-only condition than in the PT condition (mean difference = 0.081, $p = .011$) for the participants who started with the text-only condition, whereas the opposite effect occurred in the group of participants starting with the TP condition (mean difference = 0.086, $p = .005$). In the group starting with PT, there were no significant differences between conditions.

Importantly, the main effect of the experimental conditions was significant for the textbase, $F(2, 274) = 8.083$, $p < .001$, $\eta^2 = .056$, whereas no significant main effects of the experimental conditions were observed regarding the text surface, $F(2, 212) = 0.953$, $p = .387$, or the situation model, $F(2, 284) = 0.363$, $p = .696$. This suggests that the main results of our experiment were not affected by carryover effects.

Table 2

Acceptance Rates per Sentence Type, Mean Reading and Picture-Viewing Times

	Acceptance rate (relative frequency of "yes" responses) per sentence type (N = 146)				Mean reading/viewing time in milliseconds (N = 146)	
	Original	Paraphrase	Meaning Change	Situation Change	Text (without titles)	Illustrations
Mean	.862	.744	.602	.287	8685.18	2085.48
Median	.889	.750	.583	.250	8147.25	2025.54
SD	.103	.167	.179	.178	3381.01	860.19
Minimum	.444	.333	.167	.000	3408.42	870.05
Maximum	1.000	1.000	1.000	.750	25813.99	4825.05

Table 3

Mean Sensitivity A's for Surface, Textbase, and Situation Model, and Mean Reading and Picture Viewing Times (in Milliseconds) Dependent on Experimental Conditions

	Surface *A'* (N = 109)		Textbase *A'* (N = 139)		Situation model *A'* (N = 145)		Mean reading time (N = 146)		Mean picture-viewing time (N = 146)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Text-only	.593	.257	.649	.237	.833	.143	9098.31	4139.41	---	---
Text-picture	.600	.243	.579	.243	.822	.168	8517.44	3307.83	2161.33	1047.93
Picture-text	.559	.257	.681	.223	.823	.153	8431.00	3430.53	2010.00	867.28

Discussion

The purpose of our study was to examine the effect of illustrations on text surface, textbase, and situation model representations (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998) of written narrative text read by elementary and early secondary school children. The participants performed a sentence recognition task that allowed us to simultaneously measure all three levels (Fletcher & Chrysler, 1990; Nieding, 2006; Schmalhofer & Glavanov, 1986). As the participants were forced to sequentially process verbal text and illustrations, we were particularly interested in possible effects of the processing order. Therefore, each participant was presented with three versions of the sentence recognition task: one with sentences presented alone (text-only), one with sentences presented before their corresponding illustrations (TP), and one with the illustrations presented first (PT).

Situation model

Our hypothesis that situation model representations would benefit from the presence of illustrations was not supported by the data. Therefore, the stable superiority of audiovisual text to auditory text with regard to the situation model (e.g., Seger et al., 2019; Wannagat et al., 2018) does not seem to pertain to written text compared with illustrated written text. This finding can be interpreted in the context of the modality principle (Low & Sweller, 2014) which claims that pictures have a greater beneficial impact on text comprehension if two sensory channels are involved instead of one.

Nevertheless, several studies have reported a positive effect of illustrations on the comprehension of written stories (Gambrell & Jawitz, 1993; O'Keefe & Solman, 1987; Pike et al., 2010). However, there are three major differences between them and the study reported here. First, illustrations may be more beneficial when they appear together with the stories, which was the case in the studies of Gambrell and Jawitz (1993) and Pike et al. (2010). In O'Keefe and Solman's (1987) first two experiments, the advantage of stories with sequentially presented illustrations over stories without illustrations was smaller (and only significant in an analysis of both experiments combined) than the advantage of illustrations simultaneously presented with its corresponding verbal text in experiment 3. Situation model construction may benefit from features of concurrent text-picture units that are not shared by sequential ones. We cautiously assume that concurrent text-picture units provide the opportunity for the iterative processing of verbal text and pictures, which may lead to a more accurate representation of the state of affairs described.

Second, the stories used in all of these studies included fewer illustrations than ours while having a comparable (Pike et al., 2010) or even larger (Gambrell & Jawitz, 1993; O'Keefe & Solman, 1987) number of words, which results in pronouncedly different picture-per-word rates (1:15 in our study, as opposed to 1:65 in Pike et al.'s and less than 1:100 in the other two studies). If one illustration refers to a portion of text larger than 100 words, it is quite likely that this illustration helps the reader with integrating the comparably rich semantic information into a coherent situation model. By contrast, one illustration per sentence is supposed to have a more limited potential to do so; moreover, illustrations that are presented in alternation with sentences interrupt the flow of reading, which in turn may have a detrimental effect on situation model construction. Therefore, we do not rule out the possibility that illustrations would enhance situation model construction if there was only one illustration per story rather than one per sentence.

Finally, our sentence recognition task did not allow us to show pictures of inferences that were incompatible with situation change distractors. In contrast, Gambrell and Jawitz (1993) and O'Keefe and Solman (1987) tested whether the total number of correctly recalled information units differed between the illustrated and text-only conditions. Both approaches did not examine the possibility that this difference might be limited to information units that were present in both verbal text and illustrations. Moreover, the central result of the study of Pike et al. (2010) was that the generation of correct inferences was significantly enhanced when relevant features of the situation were shown. Therefore, it is possible that readers' situation models benefit from illustrations only with regard to the aspects displayed.

Textbase

Our second hypothesis was that illustrations would impede textbase representations, especially when they were presented after written text. This was confirmed insofar as textbase sensitivities were lower in the TP condition than in the other two conditions. The model inspection process, which is part of the ITPC framework (Schnotz, 2014; Schnotz & Bannert, 2003), accounts for this result: readers construct a situation model based on verbal text information and then update this model based on visual information from the illustration. This is followed by model inspection, where readers encode the updated model in a propositional format, which allows them to verbalize the story plot from their own perspective. This means that the illustrations that are presented after its corresponding verbal text can motivate readers to make substantial changes to their textbase representations.

Earlier results with auditory narrative text indicated an overall negative effect of illustrations on the textbase (Wannagat et al., 2018). The explanation was that obtaining a situation model on the depictive path of the ITPC (via analog structure mapping) would render the semantic processing of verbal text less relevant; therefore, the participants would generate a weaker text representation at the semantic level. If this was the case, we would expect lower sensitivities in both the TP and PT conditions than in the text-only condition of this study. Because textbase sensitivities were lower in TP, but not in PT, compared with text-only, this explanation would seem less suitable than the model inspection account described above. Therefore, we suggest that recipients form a textbase representation regardless of whether the text is illustrated and that illustrations can initiate model inspection, leading to changes in the textbase representation, especially when the illustration is processed after its corresponding verbal text.

Different presentation modalities of verbal text may explain why participants supposedly neglected semantic processing in Wannagat et al.'s (2018) study but not in this one. The auditory stories used by Wannagat et al. (2018) are recorded readings of written stories and do not resemble oral language, which means that textbase processing might be less effective when written text is presented aurally rather than in its original written format. By consequence, illustrations may prompt listeners to apportion less mental resources to semantic processing and to favor analog structure mapping instead, whereas readers rely more on semantic processing in the presence of illustrations because it is more effective and the beneficial effects of illustrations are weaker (modality principle). Please note that the evidence of low textbase representations in audiovisual text is scarce. On the one hand, Seger et al. (2019) did not find any effects of illustrations on textbase processing in auditory stories. On the other hand, Wannagat et al. (2017) compared written stories without illustrations with auditory and audiovisual stories in a sample of 8 and 10 year-old children as well as adults. According to the explanation outlined above, they would be expected to find higher textbase accuracies in the written than in the other two conditions or at least lower textbase accuracies in the audiovisual than in the other two conditions. However, neither of these results was obtained. To further examine this supposition, future studies should include simultaneous units of written text and illustration and compare these to audiovisual text.

Text surface

Other than we predicted, there were no significant differences between written and illustrated written text formats with respect to text surface sensitivities; this is in contrast to our earlier studies' findings that illustrations improved text surface representations of auditory text (Seger et al., 2019; Wannagat et al., 2018). Interestingly, these studies also reported a positive effect of illustrations on situation model construction. It might well be that the memory of the exact wording profits from the media features that facilitate situation model construction to the extent that the cognitive resources needed for the latter process can partly be spared when illustrations are present. This is in line with another finding reported in our 2019 study; accordingly, text surface representations were significantly improved when auditory text was furnished with static illustrations but not animated ones, whereas the situation model sensitivity was equally high in both conditions. We argued that the animations demanded additional cognitive load that used up the resources left over from the situation model construction in both audiovisual text versions. In the study reported here, neither the situation model nor the text surface profited from the presence of illustrations. At this point, however, we should be aware of the danger of over-interpreting a single non-significant result. Perhaps it is worth gauging the linear relationship between text surface and situation model representations within the scope of a systematic review or meta-analysis.

Reading and picture-viewing times

As expected, reading times were significantly shorter when illustrations were present rather than absent, corroborating the general assumption that pictures facilitate reading (i.e., multimedia principle; Mayer, 2009). However, the specific version of this assumption, namely that illustrations would diminish the reading time of *subsequent text*, could not be affirmed because there was no difference between the PT and TP conditions. One reason might be that illustrations help recipients anticipate the further course of events (i.e., support predictive inferences; cf. Unsöld & Nieding, 2009), which might constitute a reliable comprehension strategy for the commonplace stories in our study. In this case, whether the term “subsequent text” refers to the corresponding sentence (PT) or the following sentence (TP) would be of little relevance; both sentences might more easily be predicted in these conditions compared with the text-only condition, resulting in shorter reading times.

Alternatively, the participants might have been more confident about their task performance when illustrations were present and therefore spent less time reading. Although illustrations presented before or after verbal text do not seem to increase understanding, it is still possible that they increase the illusion of understanding (e.g., Jaeger & Wiley, 2014; Serra & Dunlosky, 2010). Nonetheless, the total time spent on a sentence was, on average, more than a second longer in the illustrated conditions than in the text-only condition (cf. Table 3). It can thus be stated that the ensemble of processes related to the situation model (i.e., model construction, model inspection, and analog structure mapping) in the two illustrated text versions was more time-consuming than the model construction process in the verbal-only version, without having a benefit on the situation model accuracy. We tentatively conclude that asynchronous units of written text and illustrations are inefficient media formats in the domain of narrative text (for scientific text, see also research on the temporal contiguity principle, e.g., Mayer & Fiorella, 2014).

Our participants spent significantly more time on viewing the illustrations in the TP condition than in the PT condition. We did not expect this result, but we think that it can be ascribed to the same process that may also account for the pattern of results at the textbase level; model inspection (Schnotz & Bannert, 2003) may be more pronounced when the sentence has been processed before the illustration than vice-versa. For example, imagine a participant reading "[Max] pours the sugar from the red bowl into the salt shaker." If the participant has constructed a situation model that depicts Max with the sugar bowl in his right hand and the salt shaker in his left, the subsequent illustration may induce the participant to update this situation model (cf. Figure 1) so that it depicts the sugar bowl in Max's left hand (perhaps together with the inference that Max may be left-handed). Thus, one may suppose that model inspection may take additional cognitive resources that are mirrored in longer picture-viewing times.

The differences in picture-viewing times reported here may also be caused by analogous (perceptual) simulation processes, as proposed by the embodied account of language comprehension (e.g., Glenberg & Robertson, 2000). A considerable number of studies on this topic (e.g., de Koning, Wassenburg, Bos, & van der Schoot, 2017; Engelen, Bouwmeester, de Bruin, & Zwaan, 2011; Zwaan et al., 2002) have found that after receiving a piece of text (e.g., *The ranger saw the eagle in the sky*), pictures compatible with the participant's situation model (e.g., an eagle with spread wings) are associated with shorter response times than incompatible pictures (e.g., an eagle with folded wings). Given that both pictures refer to the same linguistic concept (e.g., eagle), these results have often been interpreted in favor of the assumption that the situation model is grounded in perception and action. However, newer research attempts call into question whether these results are at all associated with perceptual simulation (Ostarek & Huettig, 2019; Ostarek, Joosen, Ishag, de Nijs, & Huettig, 2019) and whether perceptual simulation is functionally related to text comprehension (Strozyk, Dudschig, & Kaup, 2019).

Finally, mean picture-viewing times remarkably vary across participants, ranging from just below one to almost five seconds (see Table 2). Exploring in detail how students use illustrations while looking at them and how far individual differences may play a role here might be worthwhile. For example, one may imagine that those spending more time on illustrations try to create an appropriate context where the presented story may be embedded; indeed, such a strategy can support the construction of an appropriate situation model. In this sense, we encourage future research to more deeply explore what children do while they are exposed to illustrations of narrative text.

Limitations and future directions

One methodological drawback may originate from the instruction, which may have induced the participants to learn the sentences by rote (i.e., focus on the text surface) instead of constructing a situation model (i.e., "real understanding" according to Kintsch, 1998). In fact, our intention was that participants would not only focus on the situation model but also pay attention to the text surface and textbase, which were also within the scope of our research interest. In an earlier study (Author, 2019), we followed a different approach – namely, providing a rather vague instruction of *remembering the text well* and asking afterward whether the participants employed a verbatim or plot-based memory strategy. As expected, those who indicated using a verbatim strategy outperformed those employing an exclusively plot-based strategy with regard to the text surface. However, there was no such effect concerning the situation model or textbase. We inferred from these results that situation model construction unintentionally takes place during text reception at least when the text is close to the recipients' daily lives, whereas some intentional effort is required to have a more accurate memory of the wording. In the study reported here, we thus decided to formulate an instruction that also prompts participants toward memorizing the text verbatim.

The within-participant design in this study increased the statistical power of the results (compared with a between-participant design of the same sample size) and controlled for individual differences concerning reading abilities, among others. The shortcoming downside of this design was that the participants read only four stories per condition and were therefore exposed to only four paraphrases, four meaning changes, and four situation changes per condition. Thus, false alarm rates of 100% were quite likely, leading to incalculable sensitivity measures with the consequence of serious drop-out rates, especially at the text surface level. We cannot rule out the possibility that these drop-outs systematically biased our results. Additional analyses, where missing data were replaced by 0 (no sensitivity) or 0.5 (random-level sensitivity), did not reveal significantly different result patterns, indicating that there was probably no such bias. Nonetheless, we cannot evade the fact that the statistical power of our text surface results is substantially reduced. Our results further indicate that carryover effects are present between the presentation modalities in the course of the experiment. Although our main results remained stable when a group factor indicating the sequence of conditions was added, they still may account for the homogeneous results we obtained, especially with regard to the situation model.

Moreover, the internal consistencies for acceptance rates for paraphrases, meaning changes, and situation changes are below the margin of acceptability. This means that the sensitivity measures for all three levels of representation are associated with considerable measurement errors that may limit the interpretability of our results, especially if these errors are systematic (i.e., interactions between participants, measures, and experimental conditions). These low internal consistency values may be attributable to the fact that there was only one paraphrase, one meaning change, and one situation change per participant and story (as opposed to three original sentences, which yielded an acceptable internal consistency value). However, we deemed it important to have a 50% rate of correct acceptances (i.e., originals) to minimize the risk that participants do not respond significantly above chance level, the consequence being that each change type could occur only once in every six sentences.

We also admit that the text-picture units in our work do not constitute a setting that represents typical narrative reading situations for 7–13 year-old children. First, the sequential presentation of illustrations and corresponding verbal text, especially without the opportunity to "turn back pages," is far from the reality of printed or electronic books. Second, as discussed earlier, the picture-per-word rate of our stories is ten times higher than that employed by O'Keefe and Solman (1987) who used real samples from fifth-grade literature. In our study, this rate is presumably closer to what would be usual for younger children's storybooks. Third, of course, children rarely read stories in expectation of a sentence recognition task; for example, reading in the school context more often requires free retelling or cued recall. However, our major research goal was related to the examination of text surface, textbase, and situation model representations in a maximally distinct way, for which the sentence recognition task introduced by Schmalhofer and Glavanov (1986) is a well-established method. Furthermore, because we were interested in gauging the influence of illustrations on all three levels, illustrating every sentence was necessary because different sentences within a story were assigned to different probe sentence types in the task, each of them indicating success or failure with regard to different levels of representation. One of our research goals was also to determine whether the processing order of sentences and illustrations would have an impact on comprehension. We thought that the most effective way to do this would be an experimental variation of the presentation order.

Finally, a sequence of actions and utterances treated in as few as six sentences cannot easily be generalized to typical narratives in the literature of Grade 2 and higher, which exceed the length of our stories by far. Moreover, these narratives show theme progression, which we have not incorporated in our study and which our results thus cannot account for. We argue that the sentence recognition task might not be an appropriate measure for larger narrative text units that feature theme progression owing to the limited impact of single sentences on the story plot as a whole. The longer and more complex the story, the more likely it becomes that the situation changes in the same way as we used them reflect details that are not necessary for having a representation of the story's main point. Nevertheless, owing to the substantial differences between the acceptance rates of meaning and situation changes in our sentence recognition task, it can be stated that those sentences referred to as situation changes must reflect alterations at a level deeper than sentence-based meaning; notwithstanding this, we suppose that neither the story material nor the task type used in our study would be able to discriminate the situation models that merely contain the details of a story plot from those that also (or instead) have caught the main point of the story.

Therefore, we suggest applying the tripartite model to longer and more complex portions of narrative text together with a lower picture-per-word rate. Eye tracking can also be a powerful tool to not only obtain reading and picture-viewing time data in simultaneously presented text-picture units but also explore the sequence of reading and picture-viewing episodes. Both types of data can be related to outcomes relevant to situation model construction to gain a deeper understanding of the cognitive processes underlying the comprehension of illustrated narrative text. Further attempts to transfer our findings to more realistic reading situations should also ask whether an iterative processing of verbal text and pictures (e.g., the opportunity to "turn back" to the text after viewing the picture or to return to the picture after reading the text) would improve situation model construction compared with strictly sequential text-picture units or verbal text alone.

To the best of our knowledge, our study marks the first systematic attempt to establish the influence of illustrations on text surface, textbase, and situation model representations of written narrative text. It further contributes to understanding the impact of the processing order of written text and pictures on text comprehension, a topic that has been explored abundantly in the domain of expository text (Eitel & Scheiter, 2015) but scarcely in the area of narrative text. Although we do not generally think that the theories developed in the context of scientific text learning can simply be transferred to the field of narrative text comprehension, this study yields evidence that the ITPC framework originating from instructional psychology (see Schnotz & Bannert, 2003) also applies well to research on narrative text. As a practical implication, we recommend that authors and typesetters of illustrated reading books place illustrations before the corresponding text passages, as long as they want readers to remember not only the state of affairs but also the meaning that the text conveys. Meaning-based representations are supposedly relevant for some tasks in language teaching, such as re-narrations and content analyses.

References

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 206–226). New York, NY: Cambridge University Press.

Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology*, *76*, 408–417.

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

Beagles-Roos, J., & Gat, I. (1983). Specific impact of radio and television on children's story comprehension. *Journal of Educational Psychology*, *75*, 128–137.

Beentjes, J.W.J., & van der Voort, T.H.A. (1991). Children's written accounts of televised and printed stories. *Educational Technology Research & Development*, *39*(3), 15–26.

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.

Butcher, K. R. (2006). Learning from text with diagrams: promoting mental model development and inference generation. *Journal of Educational Psychology*, *98*, 182–197. https://doi.org/10.1037/0022-0663.98.1.182

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). New York, NY: Cambridge University Press.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, *20*, 155–166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: use of diagrams in computer-based training environments. *Instructional Science*, *30*, 433–464.

De Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275–277.

Duursma, E., Augustyn, M., & Zuckerman, B. (2008). Reading aloud to children: the evidence. *Archives of Disease in Childhood*, *93*, 554–557. https://doi.org/10.1136/adc.2006.106336

Eitel, A., & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. *Educational Psychology Review*, *27*, 153–180. https://doi.org/10.1007/s10648-014-9264-4

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: evidence for scaffolding. *Learning and Instruction*, *28*, 48–63. https://doi.org/10.1016/j.learninstruc.2013.05.002

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. https://doi.org/10.1016/j.jecp.2011.06.009

Engelkamp, J. (1980). Some studies on the internal structure of propositions. *Psychological Research*, *41*, 355–371.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences [Computer Software]. *Behavior Research Methods, Instruments & Computers*, *39*, 175–191.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: recognition memory for three types of textual information. *Discourse Processes*, *13*, 175–190. https://doi.org/10.1080/01638539009544752

Forster, K. I., & Forster, J. C. (2016). DMDX Version 5. Retrieved from http://www.u.arizona.edu/~kforster/dmdx/download.htm

Gambrell, L. B., & Jawitz, P. B. (1993). Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*, *28*, 264–276.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.011

Gibbons, J., Anderson, D. R., Smith, R., Field, D. E., & Fischer, C. (1986). Young children's recall and reconstruction of audio and audiovisual narratives. *Child Development*, *57*, 1014–1023.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: pictures help to build mental models. *Journal of Memory and Language*, *31*, 129–151.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379–401. https://doi.org/10.1006/jmla.2000.2714

Greenhoot, A. F., & Semb, P. A. (2008). Do illustrations enhance preschoolers' memories for stories? Age-related change in the picture facilitation effect. *Journal of Experimental Child Psychology*, *99*, 271–287. https://doi.org/10.1016/j.jecp.2007.06.005

Gunter, B., Furnham, A., & Griffiths, S. (2000). Children's memory for news: a comparison of three presentation media. *Media Psychology*, *2*, 93–118. https://doi.org/10.1207/S1532785XMEP0202_1

Guttmann, J., Levin, J. R., & Pressley, M. (1977). Pictures, partial pictures, and young children's oral prose learning. *Journal of Educational Psychology*, *69*, 473–480.

Hayes, D. S., Kelly, S. B., & Mandel, M. (1986). Media differences in children's story synopses: radio and television contrasted. *Journal of Educational Psychology*, *78*, 341–346.

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The effects of storytelling and story reading on the oral language complexity and story comprehension of young children. *Early Childhood Education Journal*, *32*, 157–163.

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning and Instruction*, *34*, 58–73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension. A Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Klein, O., & Kogan, I. (2013). Does reading to children enhance their educational success? *Child Indicators Research*, *6*, 321–344. https://doi.org/10.1007/s12187-012-9174-2

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227–246). New York, NY: Cambridge University Press.

Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educational Psychologist*, *32*, 1–19.

Mayer, R. E. (2009). *Multimedia learning*. New York, NY: Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spacial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 279–315). New York, NY: Cambridge University Press.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology*, *91*, 358–368.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, *87*, 319–334.

Nieding, G. (2006). *Wie verstehen Kinder Texte? Die Entwicklung kognitiver Repräsentationen [How do children understand text? The development of cognitive representation]*. Lengerich: Pabst.

O'Keefe, E. J., & Solman, R. T. (1987). The influence of illustrations on children's comprehension of written stories. *Journal of Reading Behavior*, *19*, 353–377.

Ostarek, M., Joosen, D., Ishag, A., de Nijs, M., & Huettig, F. (2019). Are visual processes causally involved in "perceptual simulation" effects in the sentence-picture verification task? *Cognition*, *182*, 84–94. https://doi.org/10.1016/j.cognition.2018.08.017

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398–422. https://doi.org/10.3758/bf03202613

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, *105*, 243–255. https://doi.org/10.1016/j.jecp.2009.10.006

Pressley, M. (1977). Imagery and children's learning: putting the picture in developmental perspective. *Review of Educational Research*, *47*, 585–622.

Reinwein, J. (2012). Does the modality effect exist? and if so, which modality effect? *Journal of Psycholinguistic Research*, *41*, 1–32. https://doi.org/10.1007/s10936-011-9180-4

Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology*, *94*, 138–144. https://doi.org/10.1037//0022-0663.94.1.138

Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *17*, 159–173. https://doi.org/10.1037/a0023588

Rummer, R., Schweppe, J., Fürstenberg, A., Seufert, T., & Brünken, R. (2010). Working memory inferference during processing texts and pictures: Implications for the explanation of the modality effect. *Applied Cognitive Psychology*, *24*, 164–176. https://doi.org/10.1002/acp.1546

Salomon, G., & Leigh, T. (1984). Predispositions about learning from print and television. *Journal of Communication*, *34*(2), 119–135.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: verbatim, propositional and situational representations. *Journal of Memory and Language*, *25*, 279–294.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 72–103). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Seger, B. T., Wannagat, W., & Nieding, G. (2019). How static and animated pictures contribute to multi-level mental representations of auditory text in seven-, nine-, and eleven-year-old children. *Journal of Cognition and Development*, *20*(4), 573–591. https://doi.org/10.1080/15248372.2019.1636799

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory*, *18*, 698–711. https://doi.org/10.1080/09658211.2010.506441

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*, 137–149.

Strozyk, J. V., Dudschig, C., & Kaup, B. (2019). Do I need to have my hands free to understand hand-related language? Investigating the functional relevance of experiential simulations. *Psychological Research*, *83*, 406–418. https://doi.org/10.1007/s00426-017-0900-8

Taylor, L. J., & Zwaan, R. A. (2009). Action and cognition: the case of language. *Language and Cognition*, *1*(1), 45–58. https://doi.org/10.1515/LANGCOG.2009.003

Unsöld, I. H., & Nieding, G. (2009). Die Bildung prädiktiver Inferenzen von Kindern und Erwachsenen bei der kognitiven Verarbeitung audiovisueller und auditiver Texte. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *41*(2), 87–95. https://doi.org/10.1026/0049-8637.41.2.87

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

Wannagat, W., Waizenegger, G., Hauf, J., & Nieding, G. (2018). Mental representations of the text surface, the text base, and the situation model in auditory and audiovisual texts in 7-, 9-, and 11-year-olds. *Discourse Processes*, *55*(3), 290–304. https://doi.org/10.1080/0163853X.2016.1237246

Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing*, *18*, 491–504. https://doi.org/10.1007/s10339-017-0820-y

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88. https://doi.org/10.1080/01638539909545070

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*, 229–234. https://doi.org/10.1016/j.tics.2014.02.008

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

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Appendix B: Representations of auditory text

This is the accepted manuscript version of the following research article: Seger, B. T., Wannagat, W. & Nieding, G. (2019). How static and animated pictures contribute to multi-level mental representations of auditory text in seven-, nine-, and eleven-year-old children. *Journal of Cognition and Development, 20*(4), 573—591. doi:10.1080/15248372.2019.1636799

Abstract

In our current culture, children are exposed to a huge amount of audiovisual media, of which many formats include animated pictures, such as in videos, for instance. The current study addresses the use of audiovisual media in order to increase the effectiveness of learning and teaching. We examined how auditory text, audiovisual text with static pictures, and audiovisual text with dynamic pictures affect children's text comprehension. For this purpose, we followed an approach that takes three levels of representation into account: the surface, the textbase, and the situation model. A sample of 108 children aged seven, nine, and eleven years listened to 12 narrative texts that were presented either without pictures, or with pictures that were either static or animated. We used a sentence recognition task to assess memory of all three levels of representation. Our results show that when audiovisual rather than auditory text was presented, surface and situation model representations but not textbase representations were higher. There was no difference between static and dynamic pictures for any level of representation.

Introduction

Given that more than 90 percent of children in Western countries are now regular users of audiovisual screen media (e.g., Medienpädagogischer Forschungsverbund Südwest, 2019; Rideout & Robb, 2019), investigating the processes that underlie children's mental representation of text from these media have become more and more important for developmental research. Efforts to understand these processes have already been made following the introduction of television into children's daily lives, most by comparing memory of audiovisual, auditory, and print text (Gibbons, Anderson, Smith, Field, & Fischer, 1986; Greenhoot & Semb, 2008; Gunter, Furnham, & Griffiths, 2000; Hayes, Kelly, & Mandel, 1986; Pezdek, Lehrer, & Simon, 1984). The majority of these studies found that more information was remembered when audiovisual rather than auditory media were used. However, it is sometimes not clear whether this audiovisual superiority concerns more superficial or deeper levels of information processing. In particular, the studies cited above did not take into account mental situation models, which are supposed to be crucial to children's daily life and school performance (Kintsch, 1998). It would also be of interest to ascertain whether children directly benefit from the dynamic aspect of television in addition to the effect of static pictures. We therefore investigated how seven-, nine-, and eleven-year-old children process verbal-only auditory text, auditory text with static pictures, and auditory text with animated pictures at three different levels of depth: the text surface, the textbase, and the situation model (van Dijk & Kintsch, 1983).

Three levels of text representation

Among researchers of text comprehension there is agreement that three mental representations of verbal text can be distinguished (Kintsch, 1998; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). First, the *surface* refers to the exact wording of verbal text as it is heard or read. Second, the *textbase* contains the semantic content and can be depicted as a network of propositions (Kintsch, 1988, 1998). Propositions are defined as the smallest meaning units to which a truth value can be assigned and are usually represented by predicate-argument structures (e.g., Engelkamp, 1980). For example, the sentence *Michael buys a sandwich for his son* is represented by BUY(agent: Michael; object: sandwich; beneficiary: son). When the sentence is expressed in the passive voice as *A sandwich is being bought by Michael for his son*, the same textbase is achieved with different wording. Finally, the *situation model* is a mental representation that integrates information from the text with the recipient's knowledge. In order to construct a situation model, recipients draw *inferences* that are necessary to obtain a coherent representation of a situation described in the text. While reading the sentence *Michael buys a sandwich for his son*, one may infer that Michael's son is hungry or that sandwiches are his favorite food. Beyond this, situation models can be *enriched* with inferences that are not necessary for their construction (Nieding, 2006); for example, one might imagine that Michael wears glasses.

Researchers using sentence recognition tasks have revealed all three levels of mental text representation to occur simultaneously in adults (Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986). Both sets of authors found that adults were able to discriminate between original sentences and sentences where the exact wording, but not the propositional structure, was altered, thus indicating memory for text surface. Adults were also found to discriminate between changes to the exact wording and additional changes that affected the propositional structure while remaining true to the situation (e.g., Michael buys *something to eat* for his son), thus indicating textbase representation. However, discrimination was even better for sentences that were inconsistent with the situation (e.g., Michael buys a *cookie* for his son), showing that adults use information from outside the verbal text for situation model construction. Using the same method, Nieding (2006) showed that the three-level theory was also valid for children aged between five and eleven years.

Additionally, there is evidence that memory traces of surface, textbase, and situation model are influenced by strategies of text processing: Schmalhofer and Glavanov (1986) additionally varied instructions and found that the textbase was remembered better when participants were told to summarize the text afterwards, while the situation model was remembered better when they were told to acquire knowledge.

Perceptual features of situation models

In the original work on the three-level theory (Kintsch, 1988, 1998; van Dijk & Kintsch, 1983), situation model information was assumed to be stored in an amodal and abstract manner and to be represented by predicate-argument structures (like textbase information). A different line of research, based on the embodied account of language comprehension proposed by Barsalou (1999, 2008), suggests that situation models may be grounded in perception and action (Glenberg & Robertson, 2000; Zwaan, 1999, but see also Zwaan, 2014). According to the embodied view, situation models are analogous, multidimensional, and modality-specific simulations of the situation described in language, resembling mental representations of real-life experience. Studies have shown that adults perceptually simulate various object characteristics during language comprehension, for example shape (Zwaan, Stanfield, & Yaxley, 2002), orientation (Stanfield & Zwaan, 2001), color (Richter & Zwaan, 2009), sound (Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010), and vertical movement (Kaschak, Zwaan, Aveyard, & Yaxley, 2006; Meteyard, Bahrami, & Vigliocco, 2007). There is also evidence that children perceptually simulate aspects of verbal text that are not explicitly stated, indicating that embodied cognition already occurs at an early age: Engelen, Bouwmeester, de Bruin, and Zwaan (2011) found that eight-year-old children mentally simulate the shape of objects described verbally. These findings were replicated by Hauf (2016) with six-year-olds. Motion aspects also seem to be present in children's situation models. For example, Fecica and O'Neill (2010) found that preschoolers' processing time of an auditory narrative depended on the protagonist's movement speed. Children proceeded faster through movement-related sentences when the protagonist was being driven somewhere than when she was walking there. Moreover, protagonist movements were found to be related positively to the construction of spatial situation models: Children remembered locations and related objects more accurately when they listened to a story of someone walking through these locations compared to a mere description of them (Nyhout & O'Neill, 2013) or compared to locations through which the protagonist did not walk (Barnes, Raghubar, Faulkner, & Denton, 2014; Nieding & Ohler, 1999). There is also evidence that children perceptually simulate vertical object movements (Hauf, 2016).

Influence of audiovisual presentation on children's text comprehension

Auditory versus audiovisual text

Results from comparisons between auditory and audiovisual narrative texts in children tend to favor audiovisual texts at both superficial and deep levels of representation. Children are generally found to report more inferences from audiovisual compared with auditory text (e.g., Beagles-Roos & Gat, 1983; Gibbons et al., 1986; Ricci & Beal, 2002), indicating that visual information improves situation model construction. With respect to surface and/or textbase representations, most studies have revealed that more explicit details are remembered when audiovisual rather than auditory text is presented (Beagles-Roos & Gat, 1983; Gibbons et al., 1986; Pezdek et al., 1984; Ricci & Beal, 2002). However, two studies found that more direct speech was remembered in auditory text (Beagles-Roos & Gat, 1983; Hayes et al., 1986). These empirical findings correspond to the *multimedia principle* (e.g., Butcher, 2014; Mayer, 1997), which states that learning improves when text is presented in various media formats, and to the *integrated model of text and picture comprehension* (ITPC) proposed by Schnotz (2014) explains the multimedia principle by the three-level theory. Two mechanisms of text processing are distinguished: In the first instance, perceptual surface-structure processing takes place, according to which verbal text and visual images are simultaneously coded in modality-specific sensory registers. At this stage of text processing, surface representations of verbal text are obtained in the so-called *descriptive subsystem*, whereas surface representations of visual images are obtained in the *depictive subsystem*. The next step, semantic deep-structure processing, yields a propositional representation (textbase) and a mental (situation) model. Textbase is obtained by parsing, a process that takes place within the descriptive subsystem. The situation model, in contrast, is part of the depictive subsystem. On the one hand, this means that situation model construction based on textbase representations requires the transformation of a descriptive text representation into a depictive one; on the other, that there is structural correspondence between surface representations of visual images and situation models which allows them to interact. Schnotz and Bannert (2003) used the term *structure mapping* (Gentner, 1989) to refer to this interaction between surface representations of visual images and situation model construction.

According to embodied accounts of language comprehension, situation models are based on perceptual simulations of real-life experience (Barsalou, 2008). In this way, a picture added to verbal text possibly resembles real-life experience, thereby having the potential to initiate and support perceptual simulation. Structure mapping between internal visual representations and surface representations of external pictures is therefore not restricted to structural correspondence but extends to analogical correspondence grounded in actual perception. Hence, both embodied accounts and ITPC allow for the assumption that situation model construction from auditory verbal text is more accurate when it is accompanied by appropriate pictures rather than when it is presented alone. An attempt to investigate the influence of audiovisual text on each level separately was made by Wannagat, Waizenegger, Hauf, and Nieding (2018). They used Schmalhofer and Glavanov's (1986) sentence recognition paradigm to investigate how static pictures affect surface, textbase, and situation model representations of aurally presented narratives in seven-, nine-, and eleven-year-olds. Their results indicated that static pictures support surface and situation model representations but hamper textbase representations. Wannagat et al. (2018) argued that the results concerning situation model and textbase are both predicted by ITPC: Due to structure mapping processes between the picture surface and the situation model based on verbal text, the situation model is improved whereas the path from verbal text to the situation model becomes less important, which potentially leads to a weaker textbase representation.

Static versus dynamic audiovisual text

Given that both adults and children mentally simulate movements (e.g., Barnes et al., 2014; Meteyard et al., 2007), one might assume that dynamic pictures support situation model construction from corresponding verbal text more than static ones do, simply because more relevant information can be included in structure-mapping processes. Beyond that, there is evidence that people mentally simulate actions described in language through their motor system (Glenberg & Kaschak, 2002; Zwaan & Taylor, 2006; Zwaan, Taylor, & de Boer, 2010), a process referred to as *motor resonance*. Taylor and Zwaan (2009) argue that motor resonance improves comprehension of action-related language, especially with regard to situation model construction. Specifically, adults were found to simulate actions towards and away from their body (Glenberg & Kaschak, 2002) as well as rotation behavior (Zwaan & Taylor, 2006) through their motor system. With children, it has been demonstrated that situation model construction improved when they manually re-enacted narrative text using toys (Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004) or a storyboard (Rubman & Waters, 2000). Glenberg et al. also showed that situation model construction improved when children only imagined re-enacting the narrative with toys. The authors argued that both actual and imaginary re-enactment foster mental simulation processes, which in turn support situation model construction.

Furthermore, there is neurophysiological evidence that action-related language is associated with activation in the same motor and premotor areas as those associated with overt action (Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005) and action observation (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006), particularly when these actions are related to the hands, feet, or mouth. Both findings support the assumption that animated pictures may enhance comprehension of action-related verbal text.

This study

The principal goal of this study was to improve our knowledge about how children understand narrative text that is presented via audiovisual screen media. Therefore, we aimed to investigate the influence of pictures on the comprehension of auditory narrative text in children aged seven, nine, and eleven, and to find out whether animated pictures make a difference compared with static ones. Text comprehension was conceptualized according to the three-level approach. In order to obtain separate measures for each level of representation, we used the sentence recognition method introduced by Schmalhofer and Glavanov (1986). The verbal text consisted of short narratives that reflected possible daily-life situations for children in our age range. Three versions of the text were experimentally varied: verbal-only auditory text (*auditory*), verbal auditory text with static pictures (*audiovisual-static*, AVS), and verbal auditory text with an animated version of these pictures (*audiovisual-dynamic*, AVD).

Concerning situation model representations, we predicted audiovisual text to be superior to auditory text and thus to corroborate earlier empirical findings (Wannagat, Waizenegger, & Nieding, 2017; Wannagat et al., 2018) and assumptions deduced from both embodied accounts of language (Barsalou, 2008; Glenberg & Robertson, 2000) and the ITPC (Schnotz, 2014). According to the ITPC, participants in the auditory condition would only be able to construct a situation model from textbase, whereas participants in the audiovisual conditions would be able additionally to use a picture surface representation for structure mapping. According to the embodied account, structure mapping would be based on perceptual simulation, thereby resembling real-life experience. In the auditory condition, perceptual simulation was certainly available but was restricted to internal visual images; its influence was therefore expected to be lower than that in the audiovisual conditions where it would be easier to be included in structure mapping. We further hypothesized that situation model construction from AVD text would be superior to situation model construction from AVS text, in accordance with evidence showing motor resonance of action-related verbal text and video-based observation of action to be functionally related to comprehension of action-related language (Aziz-Zadeh et al., 2006; Taylor & Zwaan, 2009; Zwaan & Taylor, 2006).

With regard to text surface, we expected that recognition would be more accurate with audiovisual than with auditory text. This would confirm earlier findings from media comparison studies (e.g., Author, 2018; Pezdek et al., 1984; Ricci & Beal, 2002) and support theoretical claims of multimedia learning research (Butcher, 2014; Mayer, 1997). We made no assumptions concerning differences between AVS and AVD at this level of representation. The ITPC claims that propositional representations are mainly based on text surface representations and should therefore not be affected by pictures. In one study, results indicated a negative effect of pictures on textbase representation (Wannagat et al., 2018), but this failed to be replicated in a later study (Wannagat et al., 2017). Hence, we did not make any assumption concerning textbase in our study.

As the text comprehension strategy turned out to be an important predictor of representation levels (Schmalhofer & Glavanov, 1986), we included it as another variable in our study. For this purpose, we instructed our participants to remember the stories *as well as possible* and afterwards asked them whether they used a verbatim or a plot-oriented strategy to perform the recognition task. Comparing strategies between subjects, we expected surface and textbase representations to be better in children who used a verbatim strategy, and situation model representation to be better in children who used a plot-oriented strategy. We also examined whether text-processing strategy was related to age or presentation mode. Being aware of potential shortcomings of self-reported memory strategies, which were addressed further in the Discussion section, we conceived this as an exploratory analysis.

Method

Participants

One hundred and eight students took part in the study, with 36 in each age group and 12 per age group and experimental condition. The youngest group consisted of second graders (19 girls, mean age = 7;9, *SD* = 0;6), the middle group of fourth graders (17 girls, mean age = 9;10, *SD* = 0;6), and the oldest group of sixth graders (18 girls, mean age = 11;9, *SD* = 0;6) from different schools in Germany. All participants were native Germans or spoke German at native-speaker level. Students only participated after their parents had signed a consent form.

Sentence recognition task

Students were given a sentence recognition task adapted from Schmalhofer and Glavanov (1986) and Wannagat et al. (2017, 2018). In this task format, subjects listened first to narrative texts. Then, after a block of four texts, they were exposed to single sentences derived from the texts and asked to decide whether they had heard each before. Sentences were presented either in their original wording, requiring a "yes" answer, or modified in one of three ways: (1) as a *paraphrase*, where the exact wording was changed without changing the meaning of the sentence, e.g., by replacing original words with synonyms or by switching from active to passive voice; (2) as a *meaning change*, where the sentence's propositional structure was modified but in a sense that was still compatible with the situation described in the text; (3) as a *situation change*, where the meaning of a word was modified in a way that was incompatible with the situation. Simply put, paraphrases are incompatible with a correct text surface representation, meaning changes are incompatible with a correct textbase representation, and situation changes are incompatible with an adequate situation model.

For the task, we developed twelve stories each containing six sentences (for an example story, see Figure 1). All the stories related to everyday events that are likely to happen in a child's life, thus requiring no domain-specific knowledge or expertise. For each original sentence, three distractors were created that met the criterion of paraphrase, meaning change, or situation change (for an example sentence, see Table 1). The task was programed using PsychoPy (Peirce, 2007), version 1.82.02, on a laptop computer. All auditory text (including instructions and a practice trial) was presented in one recorded female voice. The original sentences varied between 3.8 and 8.7 seconds in duration, followed by an inter-stimulus interval of one second during the story and two seconds at the end. Each story was preceded by an appropriate title that was also presented aurally. In the AVS condition, each sentence was accompanied by a static picture. The pictures were hand-drawn with colored pencils (see Figure 1 for an example). Each was 15*11.25 inches in size and presented synchronously with their respective sentences. We avoided including any details in the static pictures that could interfere with the distractor sentences (especially situation changes), to make sure that participants could not derive any direct advantage from the pictures. Dynamic pictures in the AVD condition were animated versions of the static ones, the amount of animation being reduced to a minimum. Our animations emphasized movements of human body parts, especially the mouth, hands, and legs, which are evidentially associated with motor activation (Aziz-Zadeh et al., 2006; see Figure 1). Animation duration was synchronized with the duration of the respective auditory text. Again, we took care that no movement was shown that could interfere with the sentence's situation change version. In the auditory condition, a white 15*11.25in screen with a black dot in the center was shown during each sentence. During the inter-stimulus intervals, the screen was black in all conditions.

no.	Original sentence	Static picture	Animation
1	Auf dem Weg zum Supermarkt bemerkt Hannah, dass sie ihren Geldbeutel verloren hat. *[On her way to the supermarket, Hannah notices that she has lost her wallet.]*		*Human body:* Hannah moving, arms and legs moving
2	Sie hat Angst, dass ihre Mutter deswegen sehr böse sein wird. *[She is afraid that her mother will be very angry for that.]*		*Human body:* right hand moving toward the face
3	So kehrt Hannah um, sucht den Gehweg ab und fängt sogar an zu weinen. *[So Hannah turns around, scans the sidewalk, and even starts to weep.]*		*Human body:* right hand moving from the mouth to the right eye, tears flowing down the face
4	Sie schaut auch unter die Büsche, aber außer viel Müll kann sie nichts finden. *[She also looks under the bushes, but she can't find anything but lots of rubbish.]*		*Human body:* head, torso, and right arm bowing slightly, left arm moving toward the bush *Other:* bush moving
5	Als sie schon aufgeben will, sieht sie auf einmal ihre Mutter mit dem Geldbeutel auf sie zurennen. *[When she already wants to give up, she suddenly finds her mother running toward her with the wallet.]*		*Human body:* Mother moving fast, arms and legs moving
6	Lächelnd erzählt ihr die Mutter, dass der Geldbeutel zu Hause auf Hannahs Tisch lag. *[Mother smiles at her and tells her that the wallet was on Hannah's table at home.]*		*Human body:* lips moving, left hand waving

Figure 1. Original sentences, static pictures, and descriptions of the animations for the text "Hannah geht Einkaufen" *[Hannah goes shopping]*. Pictures were colored in the experiment.

Table 1. Original version, paraphrase, meaning change, and situation change of the first sentence from the text “Hannah geht Einkaufen” *[Hannah goes shopping]*.

Sentence type	German original	English translation
Original	Auf dem Weg zum Supermarkt bemerkt Hannah, dass sie ihren Geldbeutel verloren hat.	On her way to the supermarket, Hannah notices that she has lost her wallet.
Paraphrase	Auf dem Weg zum Supermarkt bemerkt Hannah, dass sie ihre *Geldbörse* verloren hat.	On her way to the supermarket, Hannah notices that she has lost her *purse*.
Meaning Change	Auf dem Weg zum Supermarkt bemerkt Hannah, dass ihr Geldbeutel *nicht mehr da ist*.	On her way to the supermarket, Hannah notices that her wallet *is not there*.
Situation Change	Auf dem Weg zum Supermarkt bemerkt Hannah, dass ihr Geldbeutel *geklaut wurde*.	On her way to the supermarket, Hannah notices that her wallet *has been stolen*.

Design and procedure

The three experimental conditions (auditory, AVS, AVD) were varied between students. Assignment to experimental conditions was randomized and balanced across gender, age (within age group), and daytime (class hours) as much as possible. In all experimental conditions, stories were presented in one of twelve different orders, so that within grades and experimental groups each story was presented exactly one time at every position and one time preceded and followed by every other story.

The whole experiment consisted of three blocks of four stories each. At the beginning of each story, a piano tune (3.3s) was played, followed by the story title. Then, the whole story was presented without requiring any motor response from the participant. After the presentation of the four stories, students were aurally instructed to perform the recognition task. For this task, every story was introduced by the same piano tune and the same title. The order of the stories was the same in the task as in the presentation, only the order of the sentences within the stories was scrambled in the task. For each story, three sentences were presented aurally in the original wording, one as a paraphrase, one as a meaning change, and one as a situation change. During each sentence, the drawing of a boy with a question mark remained on the screen until the left (“no”) or right (“yes”) control button on the keyboard was pressed. Then, after one second (2s at the end of a story), the next sentence (story) was presented automatically. After accomplishing one block of four stories, students were allowed to take a short break. The whole experiment took between 24 and 30 minutes.

Students were tested individually in a vacant room in their school. They sat down in front of the laptop computer and wore headphones during the whole task. They were told that they would be listening to stories and instructed *to remember them well*. They were then presented with single sentences in an arbitrary order and asked to decide whether they had heard them before or whether they were new to them. By using the above instruction, we allowed them to choose their own strategy without pushing them towards one specific level of representation. The instruction was followed by a practice story that consisted of three sentences and three respective probes in the following order: situation change, original, paraphrase. No feedback was given at any time; after the practice story, we simply asked the students if they knew how to perform the recognition task. Whenever during a practice trial we noticed a response pattern during the practice trial revealing that participants may not have understood the instructions right (e.g., by considering the order of sentences in the task), we repeated them.

When the experiment was completed, students were asked if they had tried to keep in mind either the exact wording of the sentences or what happened in the story. The experimenter noted “verbatim” when the participant tried to remember the exact wording of the sentences, and “plot-oriented” when he or she instead tried to remember what happened in the story.

Results

The data associated with this study have been published in a repository (Seger, Wannagat, & Nieding, 2018) and are available upon reasonable request.

Acceptance rates

We considered the acceptance rates (relative frequencies of "yes" responses) for the original sentences, paraphrases, meaning changes, and situation changes to determine whether the data represented the three-level theory of text comprehension. This is the case if acceptance rates are highest for the original sentences, lower for paraphrases, still lower for meaning changes, and lowest for situation changes. An analysis of variance (ANOVA) with sentence type as within-subject factor and age group (grade) as between-subject factor revealed a significant main effect of sentence type ($F[3, 315] = 428.265$, $p < .001$). Planned comparisons indicated that acceptance of originals (.902) was higher than acceptance of paraphrases (.753), which was higher than acceptance of meaning changes (.597), which was in turn higher than acceptance of situation changes (.252, all *p*s < .001). Hence, the three-level theory is applicable to our sentence recognition task. Moreover, there was a significant main effect of age group ($F[2, 105] = 5.890$, $p = .004$), indicating that second graders were more biased towards "yes" responses than were older children (both *p*s < .001), whereas there was no difference between fourth and sixth graders ($p = .309$). Age group and sentence type did not interact significantly ($F[6, 315] = 1.130$, $p = .345$). We further checked whether compatibility with the three-level theory also pertained to each of our 12 stories. The same pattern of acceptance rates emerged in every story, confirmed by significant Friedman rank tests in every case (all $\chi^2[3] \geq 50$, all *p*s < .001). Table 2 summarizes the acceptance rates for all the sentence types in every age group.

Table 2. *Mean percentages and standard deviations of yes responses to original sentences, paraphrases, meaning changes, and situation changes for 7-, 9-, and 11-year-olds.*

	originals		paraphrases		meaning changes		situation changes	
age	*M (%)*	*SD*	*M (%)*	*SD*	*M (%)*	*SD*	*M (%)*	*SD*
7 (*N* = 36)	92.74	6.16	82.64	15.22	62.73	20.75	31.71	19.08
9 (*N* = 36)	88.89	7.39	70.36	20.74	55.09	20.43	21.30	13.44
11 (*N* = 36)	89.12	8.60	72.91	18.08	61.34	19.64	22.69	15.38

Sensitivity measures

Sensitivity in the sense of Signal Detection Theory (Stanislaw & Todorov, 1999) was computed to represent the levels of representation. Specifically, we chose *A'*, which is a nonparametric sensitivity measure that is independent of response bias (criterion) and ranges from 0 to 1, with .5 indicating that participants are responding at the chance level (Donaldson, 1992). For surface *A'*, we assigned the correct "yes" responses to originals to hits, and the "yes" responses to paraphrases to false alarms. For textbase *A'*, "yes" responses to originals and paraphrases were defined as hits, "yes" responses to meaning changes as false alarms. Finally, for situation model *A'* we categorized "yes" responses to originals, paraphrases, and meaning changes as hits, "yes" responses to situation changes as false alarms. Generally speaking, we conceived "yes" responses to a specific change type as false alarms, indicating that the subject had no adequate text representation at the corresponding level, and the combined acceptance rates at the more superficial levels as hits. *A'*s for every sentence type in every age group and condition are shown in Table 3. One-sample unilateral *t*-Tests indicate whether sensitivities are significantly above the level of chance (.50). This was the case for all *A'*s except for surface *A'*s in the auditory (grades 2 and 6) and AVD conditions (grades 2 and 4).

Our *A'* sensitivity measures can be seen as indicators of the specific contribution of one level of representation to the whole comprehension process. Note, however, that there is a linear interdependency between *A'*s because false alarms at one level are considered hits at the next deeper level. Therefore, *A'*s do not allow for direct comparisons between levels of representation and must instead be analyzed separately. Hence, we conducted univariate ANOVAs for each level of representation, with age group (2^{nd} vs. 4^{th} vs. 6^{th} grade) and presentation mode (AVD vs. AVS vs. auditory) as between-subject factors. When a significant main effect emerged, we conducted multiple post-hoc comparisons with Bonferroni-adjusted *p*-values.

Surface

The ANOVA for surface *A'* revealed a significant main effect of condition ($F[2, 98] = 3.738$, $p = .027$, $\eta^2 = .071$). Post-hoc comparisons revealed a significant advantage for AVS over auditory text ($p = .021$), thus meeting our expectation. However, there was no difference between AVD and auditory presentation ($p = .381$) or between AVD and AVS presentation ($p = .666$). The main effect of age group did not reach significance ($F[2, 98] = 1.377$, $p = .257$), nor did the interaction ($F[4, 98] = 1.026$, $p = .398$).

Textbase

For textbase *A'*, the main effect of presentation mode did not reach significance ($F[2, 99] = 1.324$, $p = .271$), neither did the main effect of age group or the interaction (*F*s < 1).

Situation model

For the situation model, there was a significant main effect of presentation mode ($F[2, 99] = 5.025$, $p = .008$, $\eta^2 = .092$). Post-hoc comparisons showed significant differences between AVD and auditory ($p = .022$) and between AVS and auditory ($p = .021$), both in the expected direction. Contrary to our hypothesis, there was no difference between the AVD and AVS conditions ($p > .999$). Interestingly, a significant main effect of age group emerged ($F[2, 99] = 3.178$, $p = .046$, $\eta^2 = .060$), with a tendency towards higher *A'*s in 2^{nd} graders compared with 4^{th} ($p = .080$) and 6^{th} graders ($p = .111$). As we had made no predictions regarding this result, we decided not to interpret it further. There was no interaction between age group and presentation mode ($F < 1$).

Table 3. *A' sensitivities for text surface, textbase, and situation model as a function of age and presentation mode. Cell size is N = 12, if not otherwise specified.*

	auditory		audiovisual-static		audiovisual-dynamic	
A'-values	*M (SD)*	90% CI	*M (SD)*	90% CI	*M (SD)*	90% CI
7-year-olds						
text surface	.54 (.17)	[.45, .62]	.64 (.22)*	[.52, .76]	.59 (.24)	[.46, .70]
textbase	.85 (.08)	[.81, .89]	.85 (.06)	[.82, .88]	.85 (.05)	[.83, .88]
situation model	.89 (.07)	[.86, .93]	.95 (.05)	[.92, .97]	.94 (.05)	[.92, .97]
9-year-olds						
text surface	.59 (.16)	[.50, .68]	.76 (.12)	[.70, .83]	.61 (.21)	[.50, .71]
textbase	.83 (.10)	[.78, .88]	.89 (.04)	[.87, .91]	.86 (.06)	[.84, .89]
situation model	.85 (.10)	[.80, .90]	.91 (.06)	[.88, .95]	.90 (.04)	[.88, .92]
11-year-olds						
text surface	.57 (.17)	[.49, .66]	.66 (.18)	[.57, .75]	.71 (.17)	[.62, .80]
textbase	.85 (.08)	[.80, .89]	.87 (.07)	[.83, .91]	.87 (.07)	[.83, .90]
situation model	.87 (.09)	[.82, .92]	.89 (.08)	[.85, .93]	.91 (.10)	[.86, .96]

CI = confidence interval. *$N = 11$.

Comprehension strategy

According to their responses when asked which strategy they used to manage their task, students were assigned to one of two groups: One group reported trying to remember the story plots without making any effort to keep the sentences verbatim in their minds ($N = 65$). The other group ($N = 43$) reported trying to remember the sentences word by word. Strategy was not significantly related to age group ($\chi^2[2] = 2.859$, $p = .265$). There was also no association with presentation mode ($\chi^2 < 1$). As expected, surface *A'*s were higher for verbatim than for plot-oriented strategies ($t[105] = 2.439$, $p = .016$). However, no significant differences between strategies could be discerned with regard to either textbase or situation model representations.

Discussion

This study was conducted to find out whether and how static and animated pictures contribute to the comprehension of narrative texts presented aurally to children aged seven, nine, and eleven. Text comprehension was defined in accordance with the three-level theory (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998), taking the text surface, the textbase, and the construction of an adequate situation model into account. To operationalize text comprehension according to this theory, we chose the sentence recognition method introduced by Schmalhofer and Glavanov (1986) and successfully tried with children (Author, 2006). This study attempts to replicate the Wannagat et al. (2017) and Wannagat et al. (2018) studies and to extend them in two ways: Firstly, we included a dynamic (animated) picture condition. Secondly, we examined whether levels of representation differed according to children's memory strategies as this seems to be the case with adults (Schmalhofer & Glavanov, 1986).

Levels of representation

Situation Model

We expected that sensitivity for the situation model (i.e., a coherent and correct representation of what happens in the story) would be higher with audiovisual than with auditory text. In our data, both AVS (static) and AVD (dynamic) conditions revealed significantly higher situation model sensitivities than the auditory condition did. This result corroborates findings from earlier media comparison studies (Beagles-Roos & Gat, 1983; Gibbons et al., 1986; Ricci & Beal, 2002; Wannagat et al., 2018) and supports our combined assumptions based on the ITPC (Schnotz, 2014) and embodied accounts of language comprehension (Barsalou, 2008). According to these assumptions, audiovisual text allows for direct interaction between an internal visual representation of a situation described verbally, and a surface representation of an external picture. This interaction is consistent with structure mapping theory (Gentner, 1989) and resembles real-life experience due to perceptual simulation processes. Hence, our results suggest that structure mapping based on pictures supports situation model representations of auditory verbal text.

Our assumption that AVD text would exhibit higher situation model sensitivities than AVS text would was not met, indicating that animations have no impact on text comprehension beyond the impact of pictures in general. This suggests that dynamic pictures do not necessarily make a particular contribution to situation model construction; nor do they support the assumption that children need motor resonance (Taylor & Zwaan, 2009) for optimal understanding of action-related verbal text. Nonetheless, alternative explanations may account for this result in our study. On the one hand, animations may not be necessary to enhance comprehension of narrative texts closely related to children's everyday life. Most of the movements described in our narratives are frequently experienced and/or observed by children in our age group (e.g., getting out of bed, running downstairs, looking at an opening door). Our participants may therefore have imagined the corresponding movements and perceptually re-experienced them without any noteworthy effort when seeing only static pictures. For that reason, it would be interesting to establish whether animations support children's situation models from narratives that are more complex and/or remote from their everyday experience.

On the other hand, the animations that we used in our study were possibly not sufficient to enhance situation model construction. This may have to do with rather technical aspects of the animations used in our study: A meta-analysis of 26 studies directly comparing the impact of static and dynamic pictures on comprehension of expository verbal text (Höffler & Leutner, 2007) found a significant advantage for dynamic pictures. This effect was moderated by the perceived realism of animations, with photo-realistic animations conveying greater advantage than less realistic ones. Hence, it is possible that enriched video material supports comprehension of auditory text more than do animations reduced to their gist. Beyond this, the effect of the animations we used may be limited because they were tagged to single sentences and did not capture the dynamic sequence of events in our narratives. Hence, other than conventional videos, our animations may have failed to aid participants in keeping track of spatial locations, protagonists, and causal structures over time, which is a crucial factor of text comprehension (Zwaan & Radvansky, 1998).

Text surface

Our assumption that pictures enhance surface representations (i.e., memory for exact wording) of auditory text was met, in line with earlier attempts showing that children remember more details from audiovisual compared with auditory narratives (Gibbons et al., 1986; Pezdek et al., 1984; Ricci & Beal, 2002; Wannagat et al., 2018). This result is consistent with the multimedia principle (Mayer, 1997), suggesting that more information from verbal text can be held in memory when it is supported by a corresponding picture. However, we obtained this effect only with static and not animated pictures. We presume that animations demand additional cognitive load (Ayres & Paas, 2007), which may have compensated for the positive effect of pictures on representations of text surface. Specifically, animations and spoken text are more transitory than static pictures, so that current information from either source must be processed while previous information is being remembered. The integration of surface information from both sources may thus have involved higher cognitive load with AVD compared with AVS text, so that surface sensitivity in the AVD condition was not significantly higher than that in the auditory condition.

Textbase

Concerning sensitivity for textbase (i.e., the representation of meaning at sentence level), there was no difference between conditions. This is compatible with ITPC in that textbase representations are mainly based on verbal text and hence should not be affected by pictures. Previous studies have revealed a negative effect of pictures on textbase (Wannagat et al., 2018)indicating that structure mapping processes may render the propositional representation of verbal text less relevant. Again, as in the study conducted by Wannagat et al. (2017), this finding was not replicated here.

Comprehension strategy

In line with our assumption, children who tried to remember the stories verbatim, at least for part of the recognition task, had better representations of text surface than did those who used a plot-oriented strategy during the whole task. By contrast, no differences between these two groups were found with regard to textbase or situation model representations. Our findings therefore suggest that different strategies of text comprehension do not affect situation model construction in children, as opposed to in adults (Schmalhofer & Glavanov, 1986). An alternative explanation, however, would be that situation model construction takes place unintentionally during text reception when the text is close to the recipient's daily life, thereby not requiring domain-specific experience. This was likely the case with our narratives, but not with the programming manuals used by Schmalhofer and Glavanov (1986), which were presented to novices in programming.

We also found that there was no association with age or presentation mode, the latter indicating that neither static nor animated pictures had an effect on children's self-reported comprehension strategies. The majority of participants in all experimental groups reported using a plot-oriented strategy during the whole task. It therefore seems that seven- to eleven-year-old children value situation model construction as the most prominent and probably the most effective way of understanding and remembering text, at least when it comes to narrative text close to their daily lives.

Limitations

Despite its potential usefulness for text comprehension research and teaching, this study comes with some limitations. One limitation concerns the size of the subsamples, which consisted of 12 participants for each age group and condition (mode of presentation). We balanced conditions across gender, age, and daytime of testing. However, we did not include further variables such as working memory updating (Pike, Barnes, & Barron, 2010) that may also affect sentence recognition performance, in order to limit the overall duration of the experiment to a reasonable amount of time. Therefore, future research with larger samples should include further variables that potentially affect sentence recognition performance or text comprehension in general.

Another potential drawback of our study is that the advantage of audiovisual over auditory text may be driven by interest. Children who saw pictures during listening may have shown greater interest, which in turn would help to improve text processing. As we did not control for text-induced interest, which is an important factor in the text comprehension literature (Hidi & Baird, 1986), we cannot rule out the possibility that interest may have a moderating effect on the influence of pictures on representations of auditory narrative text in our study. However, there is some evidence indicating that interest should not be solely responsible for our experimental results. On the one hand, interesting text has been found to improve situation model construction at the global level, whereas less interesting text was associated with better memory of the surface and local text base (McDaniel, Waddill, Finstad, & Bourg, 2000). Hence, interest may explain the advantage of audiovisual over auditory text at the situation model level, but not at the surface level of text representation in our study. On the other hand, pictures may increase interest while having no or even a negative effect on comprehension. This phenomenon is referred to as the *seductive detail effect* (e.g., Harp & Mayer, 1998). Recently, Wiley (2019) proposed that people who are shown pictures could become overconfident in their comprehension, which results in lower effort. In her study of expository text in an adult sample, comprehension deteriorated when purely decorative pictures were shown, thereby replicating the seductive detail effect. In addition, Wiley found that self-reported interest correlated positively with confidence bias. Taken together, the seemingly plausible assumption that pictures increase text comprehension by inducing interest is still lacking convincing empirical evidence. Future investigations of text-picture comprehension, especially in the domain of narrative text, should address this issue systematically.

With our openly worded instruction, we tried not to push our participants towards either surface or textbase processing in order to avoid them neglecting the situation model which is, according to Kintsch (1998), more important for daily-life and school situations than more superficial levels of representation. Nor did we explicitly want them to focus on the situation model to the possible neglect of the surface and textbase levels so that we would not even be able to measure them properly (Wannagat et al., 2018). To evaluate how our instruction was interpreted, we used self-reports to examine the strategy used by our participants to handle the task. There are two methodical points to be addressed here. First, it was difficult to achieve a good ratio of correct “yes” and “no” answers as it varied depending on levels of representation and comprehension strategies. In our experiment, the ratio would be ideal (50%) when matching correct “yes” answers to original sentences and correct “no” answers to paraphrases, meaning changes, and situation changes; say, when participants were committed to forming a memory of text surface. However, five of the six sentences (83%) required a “yes” answer when one was only concerned with the situation model, as this level of representation was not affected by paraphrases and meaning changes. Our concern was that when faced with this high proportion of required “yes” answers, it would be likely that participants would fail to give “no” answers in the few cases where they were required, if only not to disrupt the motor response tendency of pressing the “yes” button. Our finding that situation model *A'*s did not differ between verbatim and plot-oriented strategy users appears to verify this concern. However, by further analyzing the acceptance rates for the different sentence types in our data, we were able to reproduce the three-level theory in both verbatim and plot-oriented strategy users (all *p*s $< .001$), implying that all three levels of text representation can be reliably distinguished in our sample regardless of participants’ comprehension strategies.

Conversely, this means that our second concern has to be taken more seriously: that is, that levels of representation are not necessarily accessible to explicit memory, which may imply that some participants were not able to properly remember which comprehension strategy they really used. Nor do our self-reports necessarily represent the strategy participants used *a priori* but rather a strategy they were “pushed” into using by the characteristics of the task. This would be plausible if strategy changes occurred in both directions: from plot-oriented to verbatim, when noticing very subtle changes between the original and probe sentences; or from verbatim to plot-oriented when noticing that too much information had to be memorized so that it would be more efficient to reduce the amount of stored information to a minimum. We encourage future research to examine strategy use in a more sophisticated way, for example by experimentally varying instructions, not least because it might also be interesting from an educational point of view.

Conclusion

This study provides evidence that static pictures increase both text surface and situation model representations of auditory text, whereas dynamic pictures do not seem to have an additional effect on situation model construction. Future research is invited to compare the influence of static and dynamic pictures on the comprehension of more complex narrative text. One possibility might be using narratives that are more remote from children's everyday life. Perceptual features should then be less accessible to perceptual simulation so that external pictures might be even more helpful for situation model construction, possibly in a way that children might also benefit from animations. Another possibility might be a more complex plot which, for example, includes hierarchical goal structures requiring more mental resources from participants in order to maintain local and global coherence (Graesser, Singer, & Trabasso, 1994). In such cases, it may be that fewer mental resources would be left for the perceptual simulation of actions described in the text, in which case dynamic pictures might exert a greater effect on situation model construction. We also suggest using animations that are linked with text units that extend beyond the level of sentences, so that they are able to convey relevant dimensions of situation models like space, causation and protagonist goals over the course of the narration (Zwaan & Radvansky, 1998). In line with findings from instructional research taking the perceived realism of dynamic pictures into account (Höffler & Leutner, 2007), we recommend the use of photographs instead of line drawings, and live-action video sequences instead of animated line drawings.

This study may contribute significantly to the question of how children understand narrative text that is presented via audiovisual screen media. Our results provide evidence that the comprehension of auditory text is improved when visual elements are added, with regard to both the text surface and situation model representations. When applied to an educational context, our results indicate that pictures support processing on a macro-structural level, which is essential for a student to learn from text. Furthermore, pictures may offer support in instances when it is necessary to learn information by heart (e.g., commit a poem to memory). With respect to the specific contributions of dynamic visual elements, however, empirical work is still in the initial stages.

References

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: a cognitive load approach. *Applied Cognitive Psychology*, *21*, 695–700. https://doi.org/10.1002/acp.1343

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology*, *16*, 1818–1823. https://doi.org/10.1016/j.cub.2006.07.060

Barnes, M. A., Raghubar, K. P., Faulkner, H., & Denton, C. A. (2014). The construction of visual-spatial situation models in children's reading and their relation to reading comprehension. *Journal of Experimental Child Psychology*, *119*, 101–111. https://doi.org/10.1016/j.jecp.2013.10.011

Barsalou, L. W. (1999). Language comprehension: archival memory or preparation for situated action? *Discourse Processes*, *28*(1), 61–80. https://doi.org/10.1080/01638539909545069

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Beagles-Roos, J., & Gat, I. (1983). Specific impact of radio and television on children's story comprehension. *Journal of Educational Psychology*, *75*, 128–137.

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. https://doi.org/10.1016/j.actpsy.2010.06.008

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 174–205). New York, NY: Cambridge University Press.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*, 275–277.

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. https://doi.org/10.1016/j.jecp.2011.06.009

Engelkamp, J. (1980). Some studies on the internal structure of propositions. *Psychological Research*, *41*, 355–371.

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. https://doi.org/10.1016/j.cognition.2010.05.014

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: recognition memory for three types of textual information. *Discourse Processes*, *13*, 175–190. https://doi.org/10.1080/01638539009544752

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.011

Gibbons, J., Anderson, D. R., Smith, R., Field, D. E., & Fischer, C. (1986). Young children's recall and reconstruction of audio and audiovisual narratives. *Child Development*, *57*, 1014–1023.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imaged activity can enhance young children's reading ability. *Journal of Educational Psychology*, *96*(3), 424–436. https://doi.org/10.1037/0022-0663.96.3.424

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 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 and Language*, *43*, 379–401. https://doi.org/10.1006/jmla.2000.2714

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395.

Greenhoot, A. F., & Semb, P. A. (2008). Do illustrations enhance preschoolers' memories for stories? Age-related change in the picture facilitation effect. *Journal of Experimental Child Psychology*, *99*, 271–287. https://doi.org/10.1016/j.jecp.2007.06.005

Gunter, B., Furnham, A., & Griffiths, S. (2000). Children's memory for news: a comparison of three presentation media. *Media Psychology*, *2*, 93–118. https://doi.org/10.1207/S1532785XMEP0202_1

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: a theory of cognitive interest in science learning. *Journal of Educational Psychology*, *90*(3), 414–434.

Hauf, J. E. K. (2016). *Experimentelle Untersuchung der Entwicklung der kognitiven Textverarbeitung unter Berücksichtigung unterschiedlicher Darbietungsmodalitäten [Experimental investigation of the development of cognitive text processing with regard to different modalities of text presentations]* (Doctoral Dissertation). Julius-Maximilians Universität, Würzburg.

Hayes, D. S., Kelly, S. B., & Mandel, M. (1986). Media differences in children's story synopses: radio and television contrasted. *Journal of Educational Psychology*, *78*, 341–346.

Hidi, S., & Baird, W. (1986). Interestingness - a neglected variable in discourse processing. *Cognitive Science*, *10*, 179–194.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction*, *17*, 722–738. https://doi.org/10.1016/j.learninstruc.2007.09.013

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, *30*, 733–744.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension. A Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educational Psychologist*, *32*, 1–19.

McDaniel, M. A., Waddill, P. J., Finstad, K., & Bourg, T. (2000). The effects of text-based interest on attention and recall. *Journal of Educational Psychology*, *92*(3), 492–502. https://doi.org/10.1037//0022-0663.92.3.492

Medienpädagogischer Forschungsverbund Südwest (2019). KIM-Studie 2018. Kindheit, Internet, Medien. [KIM Study 2018. Childhood, Interet, Media]. Retrieved from https://www.mpfs.de/fileadmin/files/Studien/KIM/2018/KIM-Studie_2018_web.pdf

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs. Language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013.

Nieding, G. (2006). *Wie verstehen Kinder Texte? Die Entwicklung kognitiver Repräsentationen [How do children understand text? The development of cognitive representation]*. Lengerich: Pabst.

Nieding, G., & Ohler, P. (1999). Der Einfluss von Protagonisten-Zielstrukturen auf räumliche mentale Modelle beim narrativen Textverstehen von Kindern. *Sprache & Kognition*, *18*, 146–158. https://doi.org/10.1024//0253-4533.18.34.146

Nyhout, A., & O'Neill, D. K. (2013). Constructing spatial representations from narratives and non-narrative descriptions: evidence from 7-year-olds. In M. A. Finlayson, B. Fisseni, B. Löwe, & J. C. Meister (Eds.), *Workshop on computational models of narrative* (pp. 158–

165). Wadern, Germany: Leibniz-Zentrum für Informatik. https://doi.org/10.4230/OASIcs.CMN.2013.158

Peirce, J. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pezdek, K., Lehrer, A., & Simon, S. (1984). The relationship between reading and cognitive processing of television and radio. *Child Development*, *55*(6), 2072–2082.

Pike, M. M., Barnes, M. A., & Barron, R. W. (2010). The role of illustrations in children's inferential comprehension. *Journal of Experimental Child Psychology*, *105*, 243–255. https://doi.org/10.1016/j.jecp.2009.10.006

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*, 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Ricci, C. M., & Beal, C. R. (2002). The effect of interactive media on children's story memory. *Journal of Educational Psychology*, *94*, 138–144. https://doi.org/10.1037//0022-0663.94.1.138

Richter, T., & Zwaan, R. A. (2009). Processing of color words activates color representations. *Cognition*, *111*, 383–389. https://doi.org/10.1016/j.cognition.2009.02.011

Rideout, V., & Robb, M. B. (2019). *The common sense census: media use by tweens and teens*. San Francisco: Common Sense Media.

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*(3), 503–514. https://doi.org/10.1037//0022-0663.92.3.503

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: verbatim, propositional and situational representations. *Journal of Memory and Language*, *25*, 279–294.

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 72–103). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Seger, B. T., Wannagat, W., & Nieding, G. (2018). Mental representations of audiovisual text in children aged 7 to 12 [Data Set].

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*, 137–149.

Taylor, L. J., & Zwaan, R. A. (2009). Action and cognition: the case of language. *Language and Cognition*, *1*(1), 45–58. https://doi.org/10.1515/LANGCOG.2009.003

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

Wannagat, W., Waizenegger, G., Hauf, J., & Nieding, G. (2018). Mental representations of the text surface, the text base, and the situation model in auditory and audiovisual texts in 7-, 9-, and 11-year-olds. *Discourse Processes*, *55*(3), 290–304. https://doi.org/10.1080/0163853X.2016.1237246

Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing*, *18*, 491–504. https://doi.org/10.1007/s10339-017-0820-y

Wiley, J. (2019). Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Applied Cognitive Psychology*, *33*, 9–19. https://doi.org/10.1002/acp.3495

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88. https://doi.org/10.1080/01638539909545070

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*, 229–234. https://doi.org/10.1016/j.tics.2014.02.008

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

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

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 & Language*, *112*, 143–149. https://doi.org/10.1016/j.bandl.2008.11.004

Appendix C: Perceptual simulation in auditory text

This is the accepted manuscript version of the following research article: Seger, B. T., Hauf, J. E. K. & Nieding, G. (2020). Perceptual simulation of vertical object movement during comprehension of auditory and audiovisual text in children and adults. *Discourse Processes, 57*(5-6), 460-472. doi:10.1080/ 0163853X.2020.1755801

Abstract

It has been argued that people construct situation models during text reception, and that these are analogous, multimodal representations of text grounded in perception and action. On the one hand, abundant evidence has been generated that recipients perceptually simulate features of the situation described in the text. On the other hand, findings indicating that pictures facilitate situation model construction have been explained by assuming a causal link between picture processing, perceptual simulation, and situation model construction. Using a picture verification task, we tested whether five to eleven year-old children and adults perceptually simulate vertical object movements during reception of narrative text, and whether the presence or absence of pictures during text presentation makes a difference. Our results suggest that both children and adults perceptually simulate vertical object movements. In our work, perceptual simulation was not influenced by pictures, so there is no evidence that it mediates the facilitating effect of pictures on situation model construction.

Introduction

When reading or listening to text, people are assumed to form a mental model that is compatible with the state of affairs described therein. Comprehension in this sense is defined as knowing what the world would be like when the statements in the text were true (Johnson-Laird, 1983). The tripartite model of text comprehension takes three levels of text representation into account: the text surface, which refers to the exact wording, the textbase, where the explicit text information is represented in the form of propositions, and a *situation model*, which is a coherent representation of the state of affairs that includes both explicit text information and elaborated information that goes beyond the text (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). An important issue is raised about how situation models are structured: Researchers following a constructivist tradition conceptualize situation model information like textbase information (i.e., in propositional form), assuming that the semantic content acquires its meaning from the connections between these propositions (Kintsch, 1988, 1998). Perception-based features of text representation are usually not included in these approaches. Another line of research, based on Barsalou's (1999a) embodied account of language comprehension, suggests that situation models can be externally grounded in perception and action (Glenberg & Robertson, 2000; Zwaan, 1999, 2014). According to this view, the situation model is an analogous and multimodal simulation of the situation described in the text, resembling the mental representation of real-life experience.

More specifically, the embodied account of text comprehension is based on the assumption that memory, language, and thought are rooted in perceptual symbols (Barsalou, 1999b), which are internal representations of sensorimotor or introspective neural experience. Perceptual symbols are modal, meaning that they are represented in the same neural systems as the perceptual states that produced them. They can also be integrated into an organized system so that a whole object or situation – as far as included in the representation – can be re-experienced in its absence. This process, referred to as *perceptual simulation* can thereby include changes of object features over time. As perceptual symbols are defined by neural representations, their processing, including perceptual simulation, is assumed to take place routinely, without consciousness being necessary.

Perceptual simulation of static and dynamic object features

Several studies were conducted in order to determine if perceptual simulation takes place during text comprehension. In some of them, a picture verification task was used (e.g., Stanfield &

Zwaan, 2001). Participants were presented a sentence in which an object was included whereby implied spatial orientation was either horizontal ('John puts the pencil in the drawer') or vertical ('John puts the pencil in the cup'). After reading the sentence, a picture of an object was shown and participants had to decide whether this object was mentioned in the sentence or not. If participants perceptually simulated the pencil's orientation during reading, they were expected to verify the picture of a pencil faster when the orientation shown in the picture matched the orientation implied by the sentence, and slower when the orientations mismatch. The adult sample tested by Stanfield and Zwaan (2001) yielded a significant response time difference between matching and mismatching trials, indicating that adults perceptually simulate an object's orientation during text comprehension. Later experiments yielded similar results for other visual object characteristics like shape (Zwaan, Stanfield, & Yaxley, 2002), size (de Koning, Wassenburg, Bos, & van der Schoot, 2017), color (Richter & Zwaan, 2009; Therriault, Yaxley, & Zwaan, 2009; Zwaan & Pecher, 2012), number (Patson, George, & Warren, 2014), and distance (Winter & Bergen, 2012). Additionally, it is suggested that auditory object characteristics are also perceptually simulated (e.g., Brunyé, Ditman, Mahoney, Walters, & Taylor, 2010; Kaschak, Zwaan, Aveyard, & Yaxley, 2006; Winter & Bergen, 2012), indicating that perceptual simulation is multimodal.

In order to establish whether one perceptually simulates object movements, Zwaan and colleagues conducted an experiment similar to those carried out earlier examining perceptual simulation of static object features (Zwaan, Madden, Yaxley, & Aveyard, 2004): Participants listened to sentences that implied an object movement either towards ('Mary hurled the softball at you') or away from their body ('You hurled the softball at Mary'). After this, they were shown a quick sequence of two pictures with the second being slightly larger or smaller than the first, thereby leading to the impression of a movement towards or away from the observer. He or she had to decide whether the two pictures depicted the same object. Response times were significantly shorter when the movement direction described in the sentence matched the movement direction implied by the picture sequence versus when they mismatched. As in earlier experiments, Zwaan et al. (2004) emphasized that perceptual simulation of the movement described in the text had an effect on task performance without being required for the task. Perceptual simulation of movement direction was replicated in several studies (Bergen, Lindsay, Matlock, & Narayanan, 2007; Kaschak et al., 2005; 2006; Richardson, Spivey, Barsalou, & McRae, 2003), and there is evidence from eye-tracking that adults also simulate movement speed (Speed & Vigliocco, 2014).

Evidence for perceptual simulation in children

Other than with adults, only few studies on perceptual simulation have been attempted with children until now. Two picture verification studies reported perceptual simulation of shape and orientation (Engelen, Bouwmeester, de Bruin, & Zwaan, 2011) as well as size (de Koning et al., 2017) in elementary school children. Unsöld and Nieding (2009) tested if children and adults create predictive inferences about an object's shape while listening to short narrative texts (e.g., whether one creates the image of a broken window while hearing about someone who kicks a ball against it). Using a picture-naming task, they compared response times for predictive stimuli (broken window) and non-predictive ones (intact window). Significant response time differences were found in six year-old children *but not in older age groups*. On the one hand, this may imply that the perceptual simulation of static object features may already take place among 6 year-olds. On the other hand, it is interesting that the older age groups did not exhibit faster response times in compatible compared with incompatible trials. Unsöld and Nieding (2009) argued that six year-old children overrepresent perceptual information during situation model construction, whereas older children's and adults' situation models are more parsimonious, which means that they do not include inferences that are not task-relevant (Fincher-Kiefer, 2001). It is possible that older children and adults tend to exclude perceptual features from their situation models when exposed to larger bits of text like those used by Unsöld and Nieding (2009), whereas they include these features when only sentences are presented, which has been the case in most perceptual simulation studies until now.

Regarding movement simulation in children, Fecica and O'Neill (2010) found that processing speed of an auditory narrative was associated with a protagonist's movement speed in preschoolers aged between three and five. The protagonist was a child who walked somewhere (slow condition) or was being driven there (fast condition). The narrative was presented sentence by sentence, and participants proceeded per mouse click from one sentence to the next. Processing time was significantly shorter in the fast condition compared to the slow one, indicating that children perceptually simulated protagonist movement. Recently, Hauf, Nieding, and Seger (2020) demonstrated that children's situation models may not only include protagonist movements, but also object movements. In their picture verification study with 6 to 10 year-olds as well as adults, the participants listened to sentences that included an object moving up or down (e.g., 'The apple falls on the grass'). After this, the animated picture of an object appeared either in a matching (apple "drops" from above to the center of the screen) or mismatching way (apple “rises” from below to the center of the screen). Shorter response times in matching compared with mismatching trials indicated that participants seem to simulate object movements. There was no significant difference between age groups.

Effects of presentation medium on perceptual simulation

A host of studies has been conducted to compare children’s comprehension from audiovisual, auditory, and written text (e.g., Beagles-Roos & Gat, 1983; Gibbons, Anderson, Smith, Field, & Fischer, 1986; Gunter, Furnham, & Griffiths, 2000; Hayes, Kelly, & Mandel, 1986; Walma van der Molen & van der Voort, 2000; Wannagat, Waizenegger, Hauf, & Nieding, 2018; Wannagat, Waizenegger, & Nieding, 2017). Most results have shown that comprehension improved when audiovisual rather than auditory or written text was presented, particularly when comprehension measures were used related to the situation model (e.g., inference generation). This conforms to the *Integrated Model of Text and Picture Comprehension* (Schnotz, 2014; Schnotz & Bannert, 2003). Accordingly, comprehension of text-picture units activate two processing paths, one that is descriptive and the other depictive. Regarding the descriptive path, recipients first form a surface and then a meaning representation of verbal text, which in turn constitutes the basis of situation model construction (tripartite model, van Dijk & Kintsch, 1983). In terms of the depictive path, recipients initially create a picture-surface representation. Schnotz and Bannert (2003) contend that situation models are depictive in nature so that recipients can directly map picture-surface representations into situation models (analog structure mapping, Gentner, 1989). Situation model construction should therefore be easier when verbal text is accompanied by an appropriate picture rather than alone.

It is not clear, however, which role perceptual simulation may play in this context. Wannagat et al. (2018) and Seger, Wannagat, and Nieding (2019) posited that pictures are able to initiate and support perceptual simulation during text comprehension owing to common structural properties between the perceptual simulation and picture-surface representation. This, in turn, is assumed to facilitate situation model construction. However, direct empirical evidence for this is lacking because studies of perceptual simulation have usually employed either written or auditory sentences and did not vary between auditory and audiovisual text. Slight evidence supporting the assumption of an audiovisual superiority originates, again, from the study of predictive inferences: Results obtained by Unsöld and Nieding (2009) suggest that six year-old children generate predictive inferences of object forms after being exposed to a video with voice-over, but not after being read out from a storybook.

This study

One purpose of our study was to establish whether children aged 6, 8, and 10 as well as adults perceptually simulate vertical object movements in short narratives. Based on data revealing that adults and children perceptually simulate object movements (Bergen et al., 2007; de Koning et al., 2017; Engelen et al., 2011; Hauf et al., 2020), we hypothesized that the perceptual simulation of object features takes place in all age groups. As earlier investigations of perceptual simulation at the sentence level did not find age-related differences (Engelen et al., 2011; Hauf et al., 2020), we also decided not to make an assumption about them. It seems worth noting, however, that this study uses larger text units (narratives) which may come along with higher cognitive load so that older children and adults – but perhaps not younger children – may reduce cognitive processes that are supposedly not relevant for the specific task. Younger children, by contrast, might overrepresent perceptual object features like in an earlier study that also used narratives instead of sentences (Unsöld & Nieding, 2009). Therefore, we were interested in exploring possible differences between age groups.

We also wanted to determine if presentation modalities of text have an influence on perceptual simulation. For this purpose, we compared a merely auditory version of our narratives to an audiovisual version, where static pictures were added to spoken text. To the best of our knowledge, this has not yet been investigated directly. Based on research on situation model construction (Wannagat et al., 2018) and predictive inferences (Unsöld & Nieding, 2009), which both uncovered an audiovisual superiority effect at least in younger children, we assumed that perceptual simulation would be more pronounced if an audiovisual, rather than an auditory, text was presented.

In order to operationalize perceptual simulation, a picture-verification task similar to that used by Stanfield and Zwaan (2001) and Hauf et al. (2020) was designed. In critical trials, the movement of an object's picture was either compatible or incompatible with the movement of the same object in the narrative. Perceptual simulation was assumed to take place when response times were shorter to compatible versus incompatible trials.

Method

Participants

Our sample consisted of 192 participants who were either preschoolers (*N* = 48, 27 boys, mean age = 6;0, *SD* = 0;4), second graders (*N* = 48, 23 boys, mean age = 8;2, *SD* = 0;6), fourth graders (*N* = 48, 25 boys, mean age = 10;3, *SD* = 0;5), or young adults (*N* = 48, 22 men, mean age = 23;0, *SD* = 2;10). They were recruited from kindergartens, elementary schools, and a university in Germany. All participants were German natives or spoke German at a native-speaker level. Participants under 18 years only participated when their parents had previously signed a consent form. Two preschoolers (one male, one female) had to be excluded from the analysis owing to random responses for the picture-verification task. The final sample consisted of 190 participants.

Picture-verification task

In our picture-verification task adapted from Hauf et al. (2020), participants listened to short narratives that consisted of five sentences each (see Figure 1 for an example narrative). Narratives were related to possible events in children's daily lives. They were recorded by one female voice and presented sentence by sentence, with a silence interval of one second between sentences. After the last sentence, a sinus tone cue (440 Hz, 500-milliseconds duration) was played, and this was immediately followed by the animated picture of an object which successively appeared either from the top or bottom to the center of the screen, thereby creating the impression of a downward or upward movement, respectively. The animation consisted of four steps that took 125 milliseconds each so that the whole animation duration was 500 milliseconds. Participants then decided via keypress whether the depicted object had appeared in the narrative before. The picture remained on screen until the '1" (for 'no', labelled with a sad smiley) or '3" (for 'yes', labelled with a happy smiley) of the numeric keyboard on the right was pressed or after a response timeout of five seconds. Response times were recorded with clock-on set at the beginning of the animation. Then, after a silent interval of two seconds, the next narrative was presented. No feedback was provided at any time during the task.

In the audiovisual condition, hand-drawn colored pictures were presented synchronously to the spoken narrative (see Figure 1). Every narrative was accompanied by three pictures which changed after the first or second and definitely after the third sentence. During a silent interval of one second between sentences, a blank white screen appeared when there was a picture change, otherwise the picture remained. In the auditory condition, a green '+" sign appeared synchronously to the sentences, and a blank white screen during the silent interval between sentences. The screen was always blank white while playing the cue tone and in the pause interval between two narratives.

For our research purpose, trials were only of interest if they included a vertical movement of an object in the last sentence of the narrative, and if that same object was shown in the picture, hence requiring a ‘yes’ response. We called these *experimental trials* and categorized them as *compatible* if the movement described in the narrative was in the same direction as the movement that could be seen in the animation, or as *incompatible*, if these movements were in the opposite direction. Figure 2 depicts an example of a compatible trial. In total, we designed 24 experimental trials, and half of them described an upward object movement at the end of the narrative, while the other half detailed a downward one. Twelve experimental trials were compatible, the other 12 were incompatible. Compatibility and movement direction were balanced over trials so that six trials were in each compatibility*direction condition. In addition, 30 filler trials were included, with six of them positive and 24 of them negative (i.e., the picture-verification task required a ‘no’ answer). Half of the negative and none of the positive filler trials featured a vertical object movement in the last sentence of the narrative. In every positive filler trial, the target object was not mentioned in the last sentence, so participants had to listen to the whole story, not just to the last sentence. The experiment was programmed and conducted with DMDX version 5 (Forster & Forster, 2016) on a laptop computer.

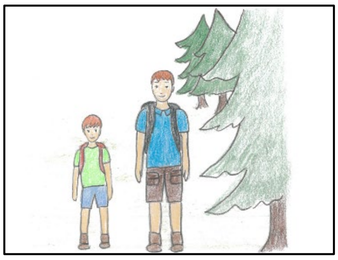
no.	Original German narrative	English translation	Picture
1	Markus und sein Vater gehen im Wald wandern.	Markus and his father go hiking in the forest.	
2	Nach einiger Zeit werden sie müde und beschließen, auf einer schattigen Bank Rast zu machen.	After a while, they're getting tired so they decide to have a break on a shady bench.	
3	Die beiden setzen sich und kramen ihr Pausenbrot aus den Rucksäcken hervor.	They sit down and dig their breaktime snacks out of their bags.	
4	Plötzlich kommt Wind auf und fegt durch die Tannen.	Suddenly, wind springs up and rips through the firs.	
5	Da fällt ein Zapfen neben Markus hinunter auf den Waldboden.	At this moment, a cone drops next to Markus on the forest floor.	

Figure 1. Sequence of sentences and pictures of an experimental narrative with an downward object movement. Pictures were colored in the experiment.

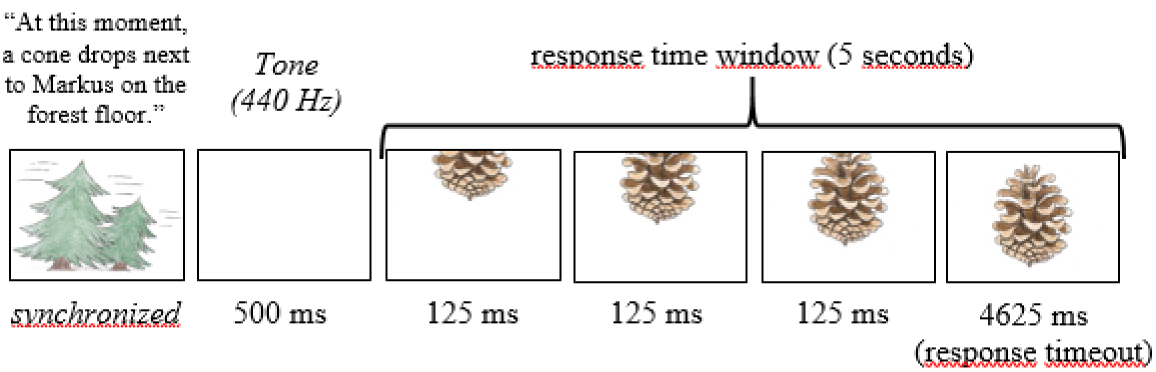


Figure 2. Time course of a compatible experimental trial (downward movement) in the audio-visual condition, beginning with the last sentence of the narrative and ending with the participant's response or response timeout. Pictures were colored in the experiment. ms = milliseconds.

Design and procedure

We employed a 2*2*2*4 design with compatibility (compatible vs. incompatible trials) and direction of the movement outlined in the story (up vs. down) as within-participant experimental factors, and modality (auditory vs. audiovisual) and age group (kindergarten vs. second grade vs. fourth grade vs. adults) as between-participant factors. Additionally, two different versions were designed so that every compatible trial in one version was incompatible in the other (and vice versa). In order to minimize sequence effects, we also varied the two versions with regard to the chronological order of trials, with one version being the inversion of the other.

Participants were tested in a vacant room at their respective educational institution (kindergarten, school, or university). They sat in front of the laptop computer and wore headphones during the entire task. They were instructed to listen carefully to the stories they were told and answer as quickly and accurately as possible during the picture-verification trials, which included placing the index and tall finger of the right hand next to the response keys so that the motor response itself would have as little influence as possible on response time. After six practice trials, the experimenter ensured that the participant understood the instructions and repeated them if necessary. Of the practice trials, three were positive and three were negative. The practice narrations did not include any vertical object movements, and no pictures were shown during these narrations. The experiment consisted of 54 trials with nine blocks containing six trials each. After each block, participants had the opportunity to take a short break. One narration was between 22 and 31 seconds in length (without the trial) and the whole experiment usually took between 30 and 35 minutes.

Data analysis

For all trials except for practice trials, accuracies and response times were recorded. Only response times to experimental items were included in further analyses, and only if the respective answers were correct. Response times exceeding five seconds (i.e., after response timeout) were counted as wrong answers and also discarded. Participants were excluded from data analysis if total accuracy was below 75% and if less than 16 response times (two-thirds of experimental items) could be considered for hypothesis testing. Unfortunately, this happened with two participants in the preschool age group.

Results

A dataset associated with this study has been published in a repository (Seger, Hauf, & Nieding, 2020) and is available from the corresponding author upon request.

Within our final sample of 190 participants, total accuracy varied between .78 and 1, with a mean accuracy of .96 and a standard deviation of .04. Mean response times to correctly answered experimental items ranged between 459 and 3234 milliseconds per participant, with an overall mean of 1407 milliseconds and a standard deviation of 550 milliseconds. For accuracy and response time data across age groups, see Table 1.

We decided to run linear-mixed model (LMM) analyses of participants' response times in order to take into account the two-level structure of our data (i.e., participants and items) since LMM allows for the computation of fixed and random effects at both levels combined (Richter, 2006; Snijders & Bosker, 2012). For the dichotomous accuracy data, we also ran a generalized LMM (GLMM) analysis. Both analyses were done via R (version 3.6.2; R Core Team, 2019) with the packages lme4 (version 3.1-1; Bates, Mächler, Bolker, & Walker, 2015) for model construction and lmerTest (version 1.1-21; Kuznetsova, Brockhoff, & Christensen, 2017) for significance testing. The MuMIn package (version 1.43.15; Barton, 2019) was used to calculate explained variances (R^2) at the model level. For the iterative procedure in the GLMM analysis, we employed a full maximum likelihood estimation (Snijders & Bosker, 2012); for the LMM, we used a restricted maximum likelihood method with generalized least square estimates, as recommended in the literature (Bates et al., 2015; Richter, 2006; Snijders & Bosker, 2012). Degrees of freedom were estimated with Satterthwaite's method (Kuznetsova et al., 2017).

The LMM for response times included the following fixed effects: compatibility (compatible vs. incompatible) as within-participant, within-item factor; modality (auditory vs. audiovisual) as between-participant, within-item factor; three contrast variables for age groups; and all two- and three-way interactions between compatibility, modality and age group. One of these age group contrasts compared preschoolers with second graders, one compared second graders with fourth graders, and one compared all three children's groups with the adult group. Random effects (intercepts) were computed at both participant and item levels. Explained variance was $R^2 = .318$ for fixed effects only and $R^2 = .659$ for fixed and random effects combined, indicating that the latter make a substantial contribution to the model. All age contrasts reached significance (preschoolers vs. second graders: $t(181.764) = 8.681$, $p < .001$; second vs. fourth graders: $t(181.350) = 6.055$, $p < .001$; children vs. adults: $t(181.659) = 9.456$, $p < .001$), and in each case, the older group's response times were shorter than those of the younger group. This likely reflects general age-related differences in processing speed (Kail & Salthouse, 1994). More importantly, compatibility was a significant predictor ($t(4170.575) = 2.735$, $p = .006$), with shorter response times being more strongly linked to compatible rather than incompatible trials. This confirmed our assumption that children and adults perceptually simulate vertical object movements. Unexpectedly, modality also appeared to be significant ($t(28.833) = 3.473$, $p < .001$), with response times for the auditory condition being notably shorter than for the audio-visual condition. However, the interaction between compatibility and modality was not significant ($t(4170.562) = 0.141$, $p = .888$), a finding which does not support our hypothesis that pictures have an impact on perceptual simulation. No other two- or three-way interaction effect was significant (all *t*s < 1). More model data, including estimates, can be found in Table 2, while Table 3 depicts response times for each compatibility and modality condition across age groups.

The GLMM for accuracy utilized the same predictors for the fixed part and the same random effects as the LMM for response times. Explained variance was $R^2 = .174$ for the fixed effects only and $R^2 = .347$ for the combined fixed and random effects, again indicating that random variance at both person and item levels contribute substantially to the model. However, no statistically significant contribution was made to the model's variance by compatibility, modality, age group, or any of the interactions between these factors (all *p*s > .05). See Table 2 for more details.

Table 1. Descriptive statistics including age (years; months), accuracy, and mean response times to experimental items across all age groups.

	Age			Accuracy			Mean response time in milliseconds		
Age group	*Mean*	*SD*	*Min. - Max.*	*Mean*	*SD*	*Min. - Max.*	*Mean*	*SD*	*Min. - Max.*
Kindergarten	6;0	0;4	5;1 - 6;8	.95	.05	.78 - 1.00	1984	435	1170 - 3233
Second grade	8;2	0;6	7;5 - 9;3	.96	.04	.83 - 1.00	1441	410	728 - 2304
Fourth grade	10;3	0;5	9;4 - 11;0	.99	.01	.94 - 1.00	1286	357	580 - 2198
Adults	23;0	2;10	19;0 - 34;6	.96	.05	.81 - 1.00	943	425	459 - 2204

Table 2. Estimates, statistics and significance levels for all significant and/or theoretically relevant predictors in the linear-mixed models for response time and accuracy.

	Linear-mixed model for response time (milliseconds)				Generalized Linear-mixed model for accuracy		
Predictor	Estimate (*b*)	*t*	*df*	*p*	Estimate (odds-ratio)	*z*	*p*
Intercept	1413.952	40.199	123.978	< .001	4.435	6.517	< .001
Age: 6 year-olds vs. 8 year-olds	412.625	8.681	181.764	< .001	0.095	0.095	.924
Age: 8 year-olds vs. 10 year-olds	284.455	6.055	181.350	< .001	1.459	0.935	.350
Age: children vs. adults	156.827	9.456	181.659	< .001	0.108	0.363	.717
Modality	100.136	3.473	181.607	< .001	0.206	0.487	.626
Compatibility	16.541	2.735	4170.575	.006	0.063	0.083	.934
Compatibility*Modality	0.854	0.141	4170.562	.888	0.296	0.625	.532

Table 3. Mean response times in milliseconds and standard deviations (in brackets) depending on age group, modality, and compatibility.

Age Group	Modality	Compatibility		
		Compatible	Incompatible	Total
6 year-olds (n = 46)	Auditory (n = 22)	1899 (500)	1933 (471)	1916 (478)
	Audiovisual (n = 24)	2024 (412)	2069 (405)	2046 (390)
	Total	1964 (455)	2004 (438)	1984 (435)
8 year-olds (n = 48)	Auditory (n = 24)	1298 (383)	1329 (367)	1315 (372)
	Audiovisual (n = 24)	1548 (387)	1585 (458)	1567 (414)
	Total	1423 (401)	1457 (431)	1441 (410)
10 year-olds (n = 48)	Auditory (n = 24)	1203 (366)	1230 (376)	1216 (368)
	Audiovisual (n = 24)	1344 (340)	1367 (357)	1355 (339)
	Total	1274 (356)	1298 (369)	1286 (357)
Adults (n = 48)	Auditory (n = 24)	782 (242)	823 (219)	802 (222)
	Audiovisual (n = 24)	1072 (550)	1108 (547)	1084 (528)
	Total	927 (445)	966 (436)	943 (425)
All participants (N = 190)	Auditory (n = 94)	1283 (545)	1316 (534)	1300 (536)
	Audiovisual (n = 96)	1497 (549)	1532 (565)	1513 (547)
	Total	1391 (556)	1425 (559)	1407 (550)

Discussion

In this study, we sought to determine whether children between six, eight, and 10 years old and adults perceptually simulate vertical object movements during comprehension of auditory narrative text, and whether this effect was moderated by the presence of pictures during text presentation. For this purpose, we designed a picture verification task adapted from Stanfield and Zwaan (2001) and Hauf et al. (2020). The results indicate that both children and adults perceptually simulate vertical object movements during text comprehension, which confirms our first hypothesis and replicates earlier results (Hauf et al., 2020). With regard to the notion that perceptual simulation is not crucial to picture verification, we suggest that it routinely takes place during comprehension, thereby supporting similar results from other studies (de Koning et al., 2017; Patson et al., 2014; Stanfield & Zwaan, 2001; Winter & Bergen, 2012; Zwaan et al., 2002; Zwaan et al., 2004; Zwaan & Pecher, 2012). Our results also indicate that perceptual simulation is associated with all of the age groups we tested, which corroborates earlier results showing that perceptual simulation was already established in children aged between seven and 13 as well as adults (de Koning et al., 2017; Engelen et al., 2011) and preschoolers (Fecica & O'Neill, 2010). Taken together, it seems that perceptual simulation is already developed by early childhood.

We also expected that perceptual simulation would be more pronounced when participants were presented with an audiovisual rather than an auditory text. Our data do not support this assumption, suggesting that audiovisual superiority, which has been found through inference generation and other measures of situation model construction (Gibbons et al., 1986; Seger et al., 2019; Unsöld & Nieding, 2009; Wannagat et al., 2018) does not necessarily apply to perceptual simulation. It is worth noting, however, that the critical object was never included in the pictures that were shown as part of the audiovisual condition, so we cannot rule out the possibility that perceptual simulation took place only for those features that were included in the pictures. Testing this assumption empirically would be a challenge due to the priming effects of having to display the target object of the picture verification task in advance. Nonetheless, there is still the possibility that including movement in the picture (i.e., using animated rather than static pictures) would support the notion of dynamic perceptual simulation in general. Unsöld and Nieding (2009) found that predictive inferences of static object features were improved by videos that included static as well as dynamic information. In the work presented here, pictures that convey only static information did not affect perceptual simulation of movement, so we cannot exclude the possibility that the simulation of dynamic object features may be affected by dynamic pictures. A future study that includes a video condition may clarify this point.

Unexpectedly, there was a significant effect of presentation mode, with participants in the audiovisual condition yielding longer response times than those in the auditory condition. We believe this reflects divergent response cue strengths between conditions: In the auditory condition, the only pictures shown were those including objects to be verified, meaning that these pictures served as a strong response cue. In the audiovisual condition, three of four pictures in the whole trial were not associated with any response, so that during the picture-verification period, participants may have taken more time to realize that they needed to provide a response.

While the majority of previous studies in this field used sentences (e.g., Engelen et al., 2011; Hauf et al., 2020; Stanfield & Zwaan, 2001; Winter & Bergen, 2012), we extended their insights to larger text units that require more information processing in general. One might argue that with more information having to be held in memory, cognitive processing needs to be more economical in order to complete the task successfully. This, in turn, means that information processing must be more restricted to task-relevant features. In our case, verifying the picture of an object did not require simulating the direction of its movement. The evidence that task-irrelevant perceptual simulation still takes place when whole narratives are presented confirms the notion that it is a highly automatic and effortless process. Moreover, if perceptual simulation could be spared when situation model construction is already supported by external pictures (see, for example, Wannagat et al., 2018), one would expect perceptual simulation to be *less* pronounced in the audiovisual compared to the auditory text condition, which was not the case here. Similarly, we also did not find any age differences, although one might expect that perceptual simulation would be less pronounced among older children and adults than among younger children because the former tend to overrepresent perceptual features during text comprehension (Unsöld & Nieding, 2009), whereas the latter tend to spare cognitive resources that are not task relevant (Fincher-Kiefer, 2001). We tentatively conclude that this adds another evidence that perceptual simulation happens automatically, whereas predictive inferences might be more resource-intensive and are supposedly not crucial to text comprehension (Graesser, Singer, & Trabasso, 1994).

However, evidence regarding the role of perceptual simulation in situation model construction remains scarce. In studies that demonstrate an advantage for audiovisual over auditory text in this respect (Seger et al., 2019; Wannagat et al., 2018), the authors reasoned that pictures initiated and supported perceptual simulation which then bolsters situation model construction. This investigation does not yield any evidence to support this assumption. Apart from the question of whether perceptual simulation facilitates the comprehension of text-picture units, there is a current debate whether it is functional for text comprehension at all (Kaup, de la Vega, Strozyk, & Dudschig, 2015; Zwaan, 2014). While certain researchers in the field of embodied cognition agree with this assumption (e.g., Barsalou, 1999b; Kaschak et al., 2005), evidence from newer research calls this into question (Strozyk, Dudschig, & Kaup, 2019). In this context, it is worth mentioning that the compatibility effect in our study is rather small, with a response time difference of only 34 milliseconds between matching and mismatching trials. A large sample size may be needed to replicate this effect, and one may also question the practical relevance of such a small difference.

Another methodological problem arises from a combination of two factors: the broad age range we examined, and the fact that we were primarily interested in response times – namely that our task had to be easy enough to ensure that six year-old participants made as few errors as possible. This task may have been associated with minimal mental load so that there was no urge to suppress task-irrelevant cognitive processes. If mental load was greater (e.g., if there was a recall task after each block of narratives), children and adults may suppress perceptual simulation in order to focus on task-relevant processing. This would be consistent with the findings of Madden and Dijkstra (2009), who related perceptual simulation to reading span in young and older adults. They found a positive relationship between the two, especially in older adults, and explained this with the notion that high-span participants had cognitive resources left for perceptual simulation, whereas low-span participants needed to concentrate more on task-relevant cognition and were therefore more likely to suppress perceptual simulation.

Finally, it would be interesting if text recipients perceptually simulate horizontal movements (i.e., from one side of the perceptual field to the other) as well as vertical ones. This question, which to the best of our knowledge has not yet been investigated, is not a trivial one. While there is considerable evidence that participants perceptually simulate horizontal object orientations while they read or listen to text (Engelen et al., 2011; Stanfield & Zwaan, 2001; Zwaan & Pecher, 2012), the evidence is less clear in the study of horizontally arranged word pairs. Here, it has been argued that semantic aspects of language processing are more relevant than embodied ones (Louwerse, 2008). It can be argued that the perceptual simulation of horizontal movements presumes that "left" and "right" can be considered embodied concepts similar to "up" and "down." Even if perceptual simulation research has not investigated this issue, there is evidence in embodiment research suggesting that this may be the case (e.g., Casasanto, 2009).

In conclusion, our results indicate that perceptual simulation is concomitant with text comprehension and that it seems to be performed with virtually no conscious effort (Barsalou, 1999b; Stanfield & Zwaan, 2001; Zwaan et al., 2004). No age differences were observed, suggesting that perceptual simulation is fully developed in six year-old children. The idea that perceptual simulation varies depending on the presence or absence of pictures was not empirically supported. Future research on embodied cognition should more directly assess its possible functionality in language comprehension, particularly with regard to situation model construction.

References

Barsalou, L. W. (1999a). Language comprehension: archival memory or preparation for situated action? *Discourse Processes*, *28*(1), 61–80. https://doi.org/10.1080/01638539909545069

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

Barton, K. (2019). MuMIn: Multi-Model Inference. R package version 1.43.15. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beagles-Roos, J., & Gat, I. (1983). Specific impact of radio and television on children's story comprehension. *Journal of Educational Psychology*, *75*, 128–137.

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.

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. https://doi.org/10.1016/j.actpsy.2010.06.008

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367. https://doi.org/10.1037/a0015854

De Koning, B. B., Wassenburg, S. I., Bos, L. T., & van der Schoot, M. (2017). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

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. https://doi.org/10.1016/j.jecp.2011.06.009

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. https://doi.org/10.1016/j.cognition.2010.05.014

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition*, *29*(2), 336–343.

Forster, K. I., & Forster, J. C. (2016). DMDX Version 5. Retrieved from http://www.u.arizona.edu/~kforster/dmdx/download.htm

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511529863.011

Gibbons, J., Anderson, D. R., Smith, R., Field, D. E., & Fischer, C. (1986). Young children's recall and reconstruction of audio and audiovisual narratives. *Child Development*, *57*, 1014–1023.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379–401. https://doi.org/10.1006/jmla.2000.2714

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395.

Gunter, B., Furnham, A., & Griffiths, S. (2000). Children's memory for news: a comparison of three presentation media. *Media Psychology*, *2*, 93–118. https://doi.org/10.1207/S1532785XMEP0202_1

Hauf, J. E. K., Nieding, G., & Seger, B. T. (2020). The development of dynamic perceptual simulations during sentence comprehension. *Cognitive Processing*, *21*(2), 197–208. https://doi.org/10.1007/s10339-020-00959-7

Hayes, D. S., Kelly, S. B., & Mandel, M. (1986). Media differences in children's story synopses: radio and television contrasted. *Journal of Educational Psychology*, *78*, 341–346.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*, 199–225.

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. https://doi.org/10.1016/j.cognition.2004.06.005

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, *30*, 733–744.

Kaup, B., de la Vega, I., Strozyk, J. V., & Dudschig, C. (2015). The role of sensorimotor processes in meaning composition. In M. H. Fischer & Y. Coello (Eds.), *Conceptual and Interactive Embodiment* (pp. 46–70). London: Routledge.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1998). *Comprehension. A Paradigm for Cognition*. Cambridge, UK: Cambridge University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effect models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*(4), 838–844. https://doi.org/10.3758/PBR.15.4.838

Madden, C. J., & Dijkstra, K. (2009). Contextual constraints in situation model construction: an investigation of age and reading span. *Aging, Neuropsychology, and Cognition*, *17*(1), 19–34. https://doi.org/10.1080/13825580902927604

Patson, N. D., George, G., & Warren, T. (2014). The conceptual representation of number. *Quarterly Journal of Experimental Psychology*, *67*(7), 1349–1365. https://doi.org/10.1080/17470218.2013.863372

R Core Team (2019). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/

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. https://doi.org/10.1016/S0364-0213(03)00064-8

Richter, T. (2006). What is wrong with ANOVA and multiple regression? Analyzing sentence reading times with hierarchical linear models. *Discourse Processes*, *41*, 221–250. https://doi.org/10.1207/s15326950dp4103_1

Richter, T., & Zwaan, R. A. (2009). Processing of color words activates color representations. *Cognition*, *111*, 383–389. https://doi.org/10.1016/j.cognition.2009.02.011

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 72–103). New York, NY: Cambridge University Press.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, *13*, 141–156. https://doi.org/10.1016/S0959-4752(02)00017-8

Seger, B. T., Hauf, J., & Nieding, G. (2020). Perceptual simulation of vertical object movements in short narratives: Picture verification data from children aged 5 to 11 and adults.

Seger, B. T., Wannagat, W., & Nieding, G. (2019). How static and animated pictures contribute to multi-level mental representations of auditory text in seven-, nine-, and eleven-year-old children. *Journal of Cognition and Development*, *20*(4), 573–591. https://doi.org/10.1080/15248372.2019.1636799

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Speed, L. J., & Vigliocco, G. (2014). Eye movements reveal the dynamic simulation of speed in language. *Cognitive Science*, *38*, 367–382. https://doi.org/10.1111/cogs.12096

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Strozyk, J. V., Dudschig, C., & Kaup, B. (2019). Do I need to have my hands free to understand hand-related language? Investigating the functional relevance of experiential simulations. *Psychological Research*, *83*, 406–418. https://doi.org/10.1007/s00426-017-0900-8

Therriault, D. J., Yaxley, R. H., & Zwaan, R. A. (2009). The role of color diagnosticity in object recognition and representation. *Cognitive Processing*, *10*, 335–342. https://doi.org/10.1007/s10339-009-0260-4

Unsöld, I. H., & Nieding, G. (2009). Die Bildung prädiktiver Inferenzen von Kindern und Erwachsenen bei der kognitiven Verarbeitung audiovisueller und auditiver Texte. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *41*(2), 87–95. https://doi.org/10.1026/0049-8637.41.2.87

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

Walma van der Molen, J. H., & van der Voort, T. H. A. (2000). The impact of television, print, and audio on children's recall of the news. *Human Communication Research*, *26*(1), 3–26.

Wannagat, W., Waizenegger, G., Hauf, J., & Nieding, G. (2018). Mental representations of the text surface, the text base, and the situation model in auditory and audiovisual texts in 7-, 9-, and 11-year-olds. *Discourse Processes*, *55*(3), 290–304. https://doi.org/10.1080/0163853X.2016.1237246

Wannagat, W., Waizenegger, G., & Nieding, G. (2017). Multi-level mental representations of written, auditory, and audiovisual text in children and adults. *Cognitive Processing*, *18*, 491–504. https://doi.org/10.1007/s10339-017-0820-y

Winter, B., & Bergen, B. (2012). Language comprehenders represent object distance both visually and auditorily. *Language and Cognition*, *4*(1), 1–16. https://doi.org/10.1515/langcog-2012-0001

Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, *28*, 81–88. https://doi.org/10.1080/01638539909545070

Zwaan, R. A. (2014). Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences*, *18*, 229–234. https://doi.org/10.1016/j.tics.2014.02.008

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. https://doi.org/10.1016/j.cogsci.2004.03.004

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: six replication attempts. *PLoS ONE*, *7*(12), e51382. https://doi.org/10.1371/journal.pone.0051382

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

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*(2), 168–171.

Appendix D: Materials of Experiments 1 and 2

Appendix D1: Stories and their illustrations

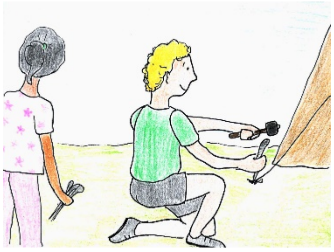
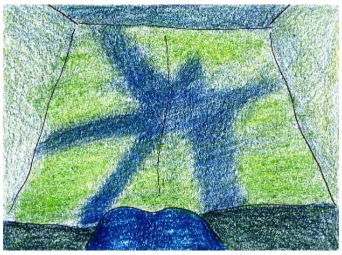
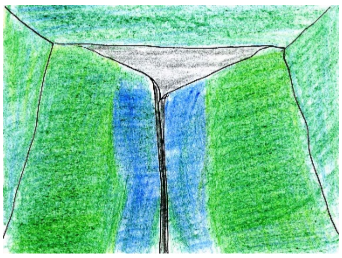
no.	Sentence	Static illustration	Animation description
1	An einem warmen Sommerabend bauen Tom, Alexa und Maja ein Zeit auf.		*Human body:* Tom hammering (left arm moving up and down); Maja approaching
2	Sie dürfen draußen übernachten und holen die Schlafsäcke und Isomatten aus dem Keller.		*Human body:* Tom, Alexa, and Maja walking into the scene (legs moving)
3	Als es dunkel wird, kuscheln sie sich in ihre Schlafsäcke und erzählen sich Gruselgeschichten.		*Human body:* Maja talking (mouth moving), Alexa's and Tom's eyes winking
4	Bald schnarchen Alexa und Maja laut, nur Tom liegt hellwach im Zelt.		*Human body:* Tom's eyes winking slowly
5	Da sieht er einen großen Schatten an der Zeltwand und kriecht tiefer in seinen Schlafsack.		*Other:* Shadow moving
6	Langsam geht der Reißverschluss des Zeltes auf und da hört Tom erleichtert die flüsternde Stimme seiner Mutter.		*Other:* Tent zip opening from above

Figure D1. Story *Die Zeltnacht* ("A Night in the Tent").

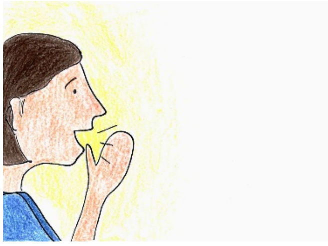
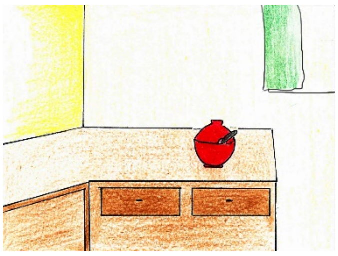
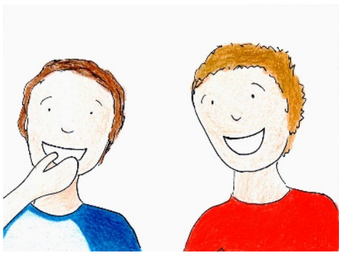
no.	Sentence	Static illustration	Animation description
1	Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade ihre Lieblingsserie anschauen.		*Human body:* Mother shouting, hand moving up
2	Die Oma kommt zum Mittagessen vorbei und Max und Sascha sollen deshalb den Tisch decken.		*Human body:* Grandmother moving hand towards her face and back
3	Als sie widerwillig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.		*Other:* Curtain moving
4	Grinsend zwinkert er Sascha zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein.		*Human body:* Left arm with sugar bowl and right arm with salt shaker moving towards each other
5	Bald sitzen alle am Tisch und Max und Sascha lassen den Salzstreuer nicht aus den Augen.		*Human body:* Brothers raising their heads and focusing their eyes on the salt shaker
6	Als der Vater dann endlich nach dem Salzstreuer greift, können die beiden ihr Lachen nicht mehr unterdrücken.		*Other:* Brothers chuckling (heads shivering), hands moving towards their mouths

Figure D2. Story *Beim Essen* (“At Lunch”).

no.	Sentence	Static illustration	Animation description
1	Auf dem Weg zum Supermarkt bemerkt Hannah, dass sie ihren Geldbeutel verloren hat.		*Human body:* Hannah moving, arms and legs moving
2	Sie hat Angst, dass ihre Mutter deswegen sehr böse sein wird.		*Human body:* right hand moving toward the face
3	So kehrt Hannah um, sucht den Gehweg ab und fängt sogar an zu weinen.		*Human body:* right hand moving from the mouth to the right eye, tears flowing down the face
4	Sie schaut auch unter die Büsche, aber außer viel Müll kann sie nichts finden.		*Human body:* head, torso, and right arm bowing slightly, left arm moving toward the bush *Other:* bush moving
5	Als sie schon aufgeben will, sieht sie auf einmal ihre Mutter mit dem Geldbeutel auf sie zurennen.		*Human body:* Mother moving fast, arms and legs moving
6	Lächelnd erzählt ihr die Mutter, dass der Geldbeutel zu Hause auf Hannahs Tisch lag.		*Human body:* lips moving, left hand waving

Figure D3. Story *Hannah geht Einkaufen* (“Hannah Goes Shopping”).

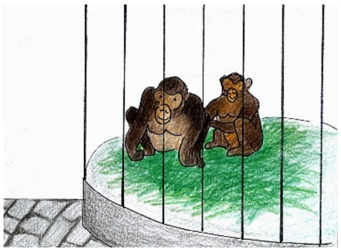
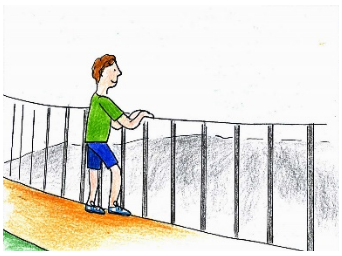
no.	Sentence	Static illustration	Animation description
1	Die Klasse macht einen Ausflug in den Zoo und Felix ist besonders auf die Elefanten gespannt.		*Human body:* Teacher and students walking towards zoo entrance
2	Zuerst gehen sie zu den Affen und beobachten fasziniert, wie die Schimpansen sich lausen.		*Other:* Right ape moving arms, left ape moving head
3	Schon bald wird Felix ungeduldig und fragt die Lehrerin, wann sie denn zu den Elefanten gehen.		*Human body:* Felix walking into scene, talking and gesturing
4	Da der Rest der Klasse aber noch die Affen fotografieren möchte, schleicht Felix sich heimlich davon.		*Human body:* Felix creeping out of scene
5	Kurz darauf steht er endlich vor dem Elefantengehege, in dem der Wärter gerade einen Babyelefanten abduscht.		*Other:* Elephant moving trunk over its head, water splashing
6	Zufrieden schaut Felix zu und hofft, dass die Lehrerin seine Flucht nicht bemerkt.		*Human body:* Felix moving head towards fence and right arm up to support head

Figure D4. Story *Im Zoo* ("At the Zoo").

no.	Sentence	Static illustration	Animation description
1	Heute feiern alle im Kindergarten Fasching und der Indianer Lukas sieht sich gespannt in der geschmückten Turnhalle um.		*Human body:* Lukas walking into scene
2	Als erstes sieht er eine gruselige Hexe und gleich daneben unterhalten sich ein Cowboy und eine Prinzessin.		*Human body:* Girl's eyes winking *Other:* Beam moving up and down
3	Er bemerkt, dass es seine drei besten Freunde sind, als sie ihm zuwinken.		*Human body:* Children waving hands
4	Schnell pirscht Lukas sich an und bewirft die drei mit einer Ladung Konfetti.		*Human body:* Lukas creeping out of scene
5	Da zieht der Cowboy seine Wasserpistole und spritzt den verdutzten Lukas nass.		*Other:* Right arm moving, left arm aligning water gun to the right
6	Lukas ergreift die Flucht und so jagen die beiden durch die Menge.		*Other:* Lukas running through scene

Figure D5. Story *Kinderfasching* ("Children's Carnival").

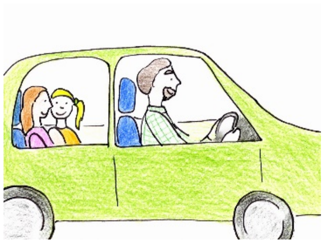
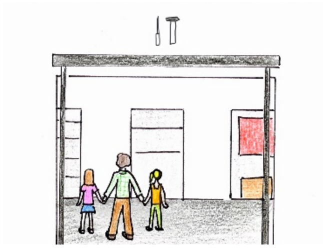
no.	Sentence	Static illustration	Animation description
1	Weil Kathi und Annika ihr Zimmer streichen dürfen, fahren sie mit ihrem Vater Farbe kaufen.		*Human body:* Daughters moving heads, father moving right hand on steering wheel *Other:* wheels turning
2	Als sie ins Geschäft gehen, sagt der Vater, dass sie sich eine Farbe aussuchen sollen.		*Human body:* Father and daughters walking through gate
3	Schon bald streiten die beiden heftig, weil Kathi lieber ein grünes und Annika lieber ein rotes Zimmer haben möchte.		*Human body:* Kathi and Annika talking, Annika gesturing
4	Bald wird der Vater wütend und droht ihnen, ohne Farbe zu gehen, wenn sie nicht gleich still sind.		*Human body:* Father talking
5	Plötzlich hat Kathi eine Idee und schnappt sich einfach zwei Farbeimer.		*Human body:* Kathi with buckets walking into scene
6	Wieder zu Hause streichen sie eine Wand grün und eine andere rot und beide Mädchen sind zufrieden.		*Human body:* Heads moving

Figure D6. Story *Die Renovierung* (“The Renovation”).

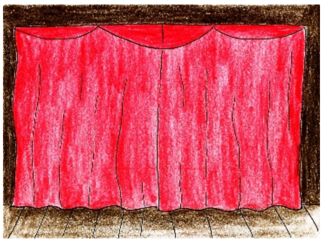
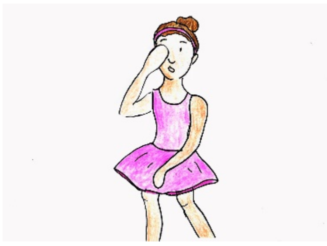
no.	Sentence	Static illustration	Animation description
1	Kurz vor dem Auftritt der Ballettgruppe ist der Vorhang noch zu und Anna und ihre Freunde laufen hinter der Bühne wild durcheinander.		*Other:* Curtain vaulting
2	Jeder geht noch einmal seine Rolle durch und alle sind sehr aufgeregt.		*Human body:* All girls except for Anna gesturing
3	Anna rennt schnell noch einmal in den Flur und sieht ein letztes Mal in den Spiegel.		*Human body:* Anna running through scene
4	Sie bemerkt, dass sich eine Haarsträhne gelöst hat und ärgert sich deswegen sehr.		*Human body:* Arms moving up, mouth and eyes frowning
5	Sie versucht zu retten, was zu retten ist und nimmt schnell einen Kamm zur Hand, aber es ist zu spät.		*Human body:* Hand (with comb) rising
6	Sie hört ihre Musik und so tritt sie wohl oder übel hinaus ins Scheinwerferlicht.		*Human body:* Right arm moving down and behind and then forth, torso moving jerkily back and forth (Anna getting into starting position)

Figure D7. Story *Anna und das Ballett* (“Anna and the Ballet”).

no.	Sentence	Static illustration	Animation description

1	Gespannt sitzen Basti und sein Vater auf der Tribüne, während sich die Fußballspieler auf dem Rasen warm machen.		*Human body:* Basti clenching right fist, father wiping with right hand over his forehead
2	Fast alle Zuschauer tragen schwarz-gelbe Schals und auch Basti bindet sich den Schal seiner Lieblingsmannschaft um.		*Human body:* Basti's left hand with scarf moving down (just finishing putting on his scarf)
3	Als er sich genauer umschaut, entdeckt er nur einige Fans der anderen Mannschaft.		*Human body:* Eyes and left hand moving to the right and back
4	Vom lauten Anfeuern bekommt er einen ganz trockenen Hals und so reicht ihm der Vater einen Becher Saft.		*Human body:* Basti shouting and raising arms
5	Auch der Vater hat schon einen ganz roten Kopf, weil heute so ein wichtiger Tag ist.		*Human body:* Father shouting
6	Der Anpfiff ertönt und Basti drückt seiner Mannschaft ganz fest die Daumen.		*Human body:* Basti moving arms up towards his chin

Figure D8. Story *Im Stadion* ("In the Stadium").

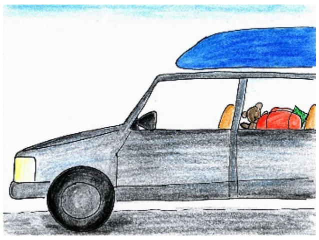
no.	Sentence	Static illustration	Animation description
1	Morgen ganz früh soll es für Familie Schmidt endlich mit dem vollgepackten Auto in den Urlaub gehen.		*Human body:* Father walking out of scene
2	Als Florian die letzten Koffer verlädt, entdeckt er überrascht, dass der rechte Hinterreifen des Autos ein Loch hat.		*Human body:* Florian approaching, lifting brown suitcase on the floor with left hand, then dropping it again
3	Er ist ganz traurig, weil er sich schon so arg auf den Urlaub gefreut hat.		*Human body:* Mouth and eyebrows getting frowny
4	Schnell holt Florian seinen Vater, damit der den Reifen reparieren kann.		*Human body:* Florian and his father walking into scene, Florian raising his left arm
5	Der Vater hebt mit seinem Wagenheber das Auto hoch und wechselt den Reifen.		*Human body:* Father drawing arms towards body *Other:* Jack lifting car
6	Damit kann morgen der Urlaub losgehen und Florian ist sehr glücklich.		*Human body:* Mouth opening, eyebrows rising

Figure D9. Story *Reisevorbereitung mit Hindernissen* ("Travel Preparation with Obstacles").

no.	Sentence	Static illustration	Animation description
1	Sandra öffnet ihrer Tante die Haustür und die beiden umarmen sich freudig.		*Human body:* Sandra opening door
2	Die Tante trägt einen geheimnisvollen Korb unterm Arm und Sandra versucht hineinzuschauen.		*Human body:* Sandra approaching basket, left arm moving behind the basket
3	Auf einmal bewegt sich der Korb und Sandra springt erschrocken zurück.		*Human body:* Sandra jumping back and raising arms
4	Da hört sie ein Winseln aus dem Korb und ein kleiner Hund streckt seinen Kopf heraus.		*Other:* Dog raising head
5	Sandra streckt ihm vorsichtig ihre Hand entgegen und der kleine Hund schleckt die Hand zutraulich ab.		*Human body:* Hand approaching dog *Other:* Dog's tongue moving
6	Sie hebt ihn aus dem Korb und Sandra tollt gemeinsam mit dem Hund im Garten herum.		*Human body:* Sandra walking into scene *Other:* Dog jumping into scene

Figure D10. Story *Besuch für Sandra* ("Visitors for Sandra").

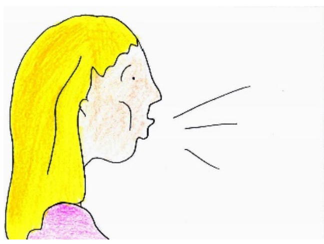
no.	Sentence	Static illustration	Animation description
1	Schlaflos wälzt sich Marie in ihrem Bett, denn morgen wird sie zehn Jahre alt.		*Human body:* Arms and right leg moving on the mattress
2	Sie denkt an die vielen Geschenke und die lieben Freunde, die zu Kakao und Kuchen kommen.		*Other:* Dishes turning clockwise around the cake
3	Um sechs hält sie es in ihrem Bett nicht mehr aus und sie steht auf.		*Human body:* Torso raising from bed with right hand used as support
4	Sie zieht gerade ihr Lieblingskleid aus dem Schrank, als sie auf dem Flur leise Stimmen hört.		*Human body:* Marie opening wardrobe door and moving right hand towards the blue dress
5	Da öffnet sich ihre Zimmertür und ihre Eltern kommen mit einer sehr großen Torte herein.		*Other:* Door opening
6	Marie bläst alle Kerzen auf Anhieb aus und wünscht sich ganz fest, dass sie jeden Tag Geburtstag hat.		*Other:* Lips forming to blow

Figure D11. Story *Maries Geburtstag* (“Marie’s Birthday”).

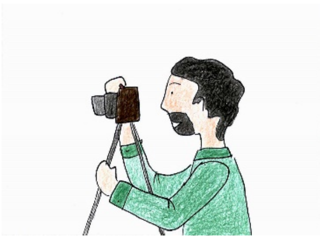
no.	Sentence	Static illustration	Animation description
1	Heute wird ein Klassenfoto gemacht und so hüpft Simon die Stufen hinunter und rennt auf den Schulhof.		*Human body:* Simon running downstairs
2	Da steht nämlich schon seine ganze Klasse bereit und wartet nur auf ihn.		*Human body:* Teacher and two students gesturing
3	Simon entschuldigt sich bei der Lehrerin und stellt sich schnell neben Martin in die erste Reihe.		*Human body:* Simon talking and gesturing, teacher bowing slightly towards Simon
4	Doch da beschwert sich die Nervensäge Sarah, weil sie hinter Simon gar nicht zu sehen ist.		*Human body:* Sarah talking and crossing arms
5	Wie immer muss Simon also in die hinterste Reihe und ist deshalb sehr sauer.		*Human body:* Simon's face turning red, mouth frowning
6	Er ist darum auch der einzige, der nicht lächelt, als der Fotograf dann das Foto macht.		*Human body:* Head bowing towards camera *Other:* Flash (picture masked white for a short time)

Figure D12. Story *Simon kommt zu spät* ("Simon's Too Late").

Appendix D2: Target sentences

Table D1

Original sentences, paraphrases, meaning changes, and situation changes of the story Die Zeltnacht ("A Night in the Tent")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	An einem warmen Sommerabend bauen Tom, Alexa und Maja ein Zelt im Garten auf.	An einem warmen Abend im Sommer bauen Tom, Alexa und Maja im Garten ein Zelt auf.	An einem schönen Sommerabend bauen Tom, Alexa und Maja ein Zelt im Garten auf.	An einem kühlen Sommerabend bauen Tom, Alexa und Maja ein Zelt im Garten auf.
2	Sie dürfen draußen übernachten und holen die Schlafsäcke und Isomatten aus dem Keller.	Sie haben die Erlaubnis, draußen zu übernachten und holen die Isomatten und Schlafsäcke aus dem Keller.	Sie dürfen im Zelt übernachten und holen die Schlafsäcke und Isomatten aus dem Keller.	Sie dürfen draußen übernachten und holen die Schlafsäcke und Isomatten vom Dachboden.
3	Als es dunkel wird, kuscheln sie sich in ihre Schlafsäcke und erzählen sich Gruselgeschichten.	Als es dunkel wird, kuscheln sie sich in ihre Schlafsäcke und erzählen sich gegenseitig gruselige Geschichten.	Als es kühler wird, kuscheln sie sich in ihre Schlafsäcke und erzählen sich Gruselgeschichten.	Als es dunkel wird, kuscheln sie sich in ihre Schlafsäcke und erzählen sich Witze.
4	Bald schnarchen Alexa und Maja laut, nur Tom liegt hellwach im Zelt.	Kurz darauf schnarchen Alexa und Maja lautstark, nur Tom liegt hellwach im Zelt.	Bald schlafen Alexa und Maja tief, nur Tom liegt hellwach im Zelt.	Bald streiten Alexa und Maja laut und Tom liegt hellwach im Zelt.
5	Da sieht er einen großen Schatten an der Zeltwand und kriecht tiefer in seinen Schlafsack.	Da sieht er einen großen Schatten an der Zeltwand und rutscht tiefer in seinen Schlafsack.	Da sieht er einen großen Schatten an der Zeltwand und versteckt sich in seinem Schlafsack.	Da sieht er einen großen Schatten an der Zeltwand und schlüpft aus seinem Schlafsack.
6	Langsam geht der Reißverschluss des Zeltes auf und da hört Tom erleichtert die flüsternde Stimme seiner Mutter.	Langsam wird der Reißverschluss des Zeltes aufgemacht und da hört Tom erleichtert die flüsternde Stimme seiner Mutter.	Langsam geht der Reißverschluss des Zeltes auf und da hört Tom erleichtert die vertraute Stimme seiner Mutter.	Langsam geht der Reißverschluss des Zeltes auf und da hört Tom erleichtert die laute Stimme seiner Mutter.

Table D2

Original sentences, paraphrases, meaning changes, and situation changes of the story Beim Essen ("At Lunch")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade ihre Lieblingsserie anschauen.	Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade ihre Lieblingsserie ankucken.	Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade lieber fernsehen.	Die Mutter ruft nach Max und Sascha, doch eigentlich wollen die beiden gerade ihr Lieblingsbuch anschauen.
2	Die Oma kommt zum Mittagessen vorbei und Max und Sascha sollen deshalb den Tisch decken.	Die Oma kommt zum Mittagessen vorbei und Max und Sascha müssen deshalb den Mittagstisch decken.	Die Oma kommt zum Mittagessen vorbei und die beiden sollen deshalb mithelfen.	Die Oma kommt zum Frühstück vorbei und Max und Sascha sollen deshalb den Tisch decken.
3	Als sie widerwillig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	Als sie widerstrebend Teller und Gläser aus der Küche herbeibringen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	Als sie widerwillig Geschirr aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.	Als sie eifrig Teller und Gläser aus der Küche herbeitragen, sieht Max auf einmal eine rote Zuckerdose auf dem Schränkchen stehen.
4	Grinsend zwinkert er Sascha zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein.	Grinsend zwinkert er seinem Bruder zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein.	Begeistert zwinkert er Sascha zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein.	Grinsend winkt er Sascha zu und schüttet dann den Zucker aus der roten Dose in den Salzstreuer hinein.
5	Bald sitzen alle am Tisch und Max und Sascha lassen den Salzstreuer nicht aus den Augen.	Bald sitzt jeder am Esstisch und Max und Sascha lassen den Salzstreuer nicht aus den Augen.	Bald sitzen alle beim Essen und Max und Sascha lassen den Salzstreuer nicht aus den Augen.	Bald stehen alle am Buffet und Max und Sascha lassen den Salzstreuer nicht aus den Augen.
6	Als der Vater dann endlich nach dem Salzstreuer greift, können die beiden ihr Lachen nicht mehr unterdrücken.	Als der Vater dann endlich den Salzstreuer nimmt, können die beiden ihr Lachen nicht mehr unterdrücken.	Als der Vater dann endlich nachsalzen will, können die beiden ihr Lachen nicht mehr unterdrücken.	Als die Oma dann endlich nach dem Salzstreuer greift, können die beiden ihr Lachen nicht mehr unterdrücken.

Table D3

Original sentences, paraphrases, meaning changes, and situation changes of the story Hannah geht Einkaufen ("Hannah Goes Shopping")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Auf dem Weg zum Supermarkt bemerkt Hannah, dass sie ihren Geldbeutel verloren hat.	Auf dem Weg zum Supermarkt bemerkt Hannah, dass sie ihre Geldbörse verloren hat.	Auf dem Weg zum Supermarkt bemerkt Hannah, dass ihr Geld-beutel nicht mehr da ist.	Auf dem Weg zum Supermarkt bemerkt Hannah, dass ihr Geld-beutel geklaut wurde.
2	Sie hat Angst, dass ihre Mutter deswegen sehr böse sein wird.	Sie hat Angst, dass ihre Mutter darum sehr wütend sein wird.	Sie hat Angst, dass ihre Mutter deswegen sehr schimpfen wird.	Sie hat Angst, dass ihre Mutter deswegen sehr traurig sein wird.
3	So kehrt Hannah um, sucht den Gehweg ab und fängt sogar an zu weinen.	So dreht Hannah um, sucht den Bürgersteig ab und beginnt so-gar zu weinen.	So kehrt Hannah um, schaut ge-nau auf den Gehweg und fängt sogar an zu weinen.	So kehrt Hannah um, sucht den Parkplatz ab und fängt sogar an zu weinen.
4	Sie schaut auch unter die Bü-sche, aber außer viel Müll kann sie nichts finden.	Sie schaut auch unter die Bü-sche, aber außer viel Abfall kann sie nichts entdecken.	Sie schaut auch unter die Bü-sche, aber den Geldbeutel kann sie nicht finden.	Sie schaut auch unter die Bü-sche, aber außer viel Löwen-zahn kann sie nichts finden.
5	Als sie schon aufgeben will, sieht sie auf einmal ihre Mutter mit dem Geldbeutel auf sie zu rennen.	Als sie bereits aufgeben möchte, sieht sie auf einmal ihre Mutter mit dem Geldbeutel auf sie zu rennen.	Als sie schon nicht mehr weiter-suchen will, sieht sie auf einmal ihre Mutter mit dem Geldbeutel auf sie zu rennen.	Als sie schon ausrasten will, sieht sie auf einmal ihre Mutter mit dem Geldbeutel auf sie zu rennen.
6	Lächelnd erzählt ihr die Mutter, dass der Geldbeutel zu Hause auf Hannahs Tisch lag.	Lächelnd erzählt ihr die Mutter, dass der Geldbeutel daheim auf Hannahs Tisch lag.	Lächelnd erzählt ihr die Mutter, dass der Geldbeutel zu Hause in Hannahs Zimmer lag.	Lächelnd erzählt ihr die Mutter, dass der Geldbeutel zu Hause unter Hannahs Tisch lag.

Table D4

Original sentences, paraphrases, meaning changes, and situation changes of the story Im Zoo ("At the Zoo")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Die Klasse macht einen Ausflug in den Zoo und Felix ist besonders auf die Elefanten gespannt.	Die Schulklasse unternimmt einen Ausflug in den Zoo und Felix ist vor allem auf die Elefanten gespannt.	Die Klasse macht einen Ausflug in den Zoo und Felix freut sich besonders auf die Elefanten.	Die Klasse macht einen Ausflug in den Zoo und alle sind besonders auf die Elefanten gespannt.
2	Zuerst gehen sie zu den Affen und beobachten fasziniert, wie die Schimpansen sich lausen.	Zunächst gehen sie zu den Affen und beobachten begeistert, wie die Schimpansen sich lausen.	Zuerst schauen sie sich die Affen an und beobachten fasziniert, wie die Schimpansen sich lausen.	Zuerst fahren sie zu den Affen und beobachten fasziniert, wie die Schimpansen sich lausen.
3	Schon bald wird Felix ungeduldig und fragt die Lehrerin, wann sie denn zu den Elefanten gehen.	Bald schon wird Felix unruhig und fragt die Lehrerin, wann sie denn zu den Elefanten gehen.	Schon bald wird Felix ungeduldig und fragt die Lehrerin, wann sie denn weitergehen.	Schon bald wird Felix müde und fragt die Lehrerin, wann sie denn zu den Elefanten gehen.
4	Da der Rest der Klasse aber noch die Affen fotografieren möchte, schleicht Felix sich heimlich davon.	Da die übrige Klasse jedoch noch die Affen fotografieren möchte, schleicht Felix sich heimlich davon.	Da der Rest der Klasse aber noch bei den Affen bleiben möchte, schleicht Felix sich heimlich davon.	Da der Rest der Klasse aber noch die Affen füttern möchte, schleicht Felix sich heimlich davon.
5	Kurz darauf steht er endlich vor dem Elefantengehege, in dem der Wärter gerade einen Babyelefanten abduscht.	Kurze Zeit später ist er endlich vor dem Elefantengehege, in dem der Wärter gerade einen Babyelefanten abduscht.	Kurz darauf steht er vor dem lange erwarteten Elefantengehege, in dem der Wärter gerade einen Babyelefanten abduscht.	Nach langem Suchen steht er endlich vor dem Elefantengehege, in dem der Wärter gerade einen Babyelefanten abduscht.
6	Zufrieden schaut Felix zu und hofft, dass die Lehrerin seine Flucht nicht bemerkt.	Zufrieden schaut Felix zu und hofft, dass seine Flucht von der Lehrerin nicht bemerkt wird.	Zufrieden schaut Felix zu und hofft, dass die Lehrerin sein Fehlen nicht bemerkt.	Zufrieden schaut Felix zu und hofft, dass die Klassenkameraden seine Flucht nicht bemerken.

Table D5

Original sentences, paraphrases, meaning changes, and situation changes of the story Kinderfasching ("Children's Carnival")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Heute feiern alle im Kindergarten Fasching und der Indianer Lukas sieht sich gespannt in der geschmückten Turnhalle um.	An diesem Tag feiern alle im Kindergarten Fasching und der Indianer Lukas sieht sich gespannt in der dekorierten Turnhalle um.	Heute feiern alle im Kindergarten Fasching und der Indianer Lukas sieht sich gespannt in der kunterbunten Turnhalle um.	Heute feiern alle in der Schule Fasching und der Indianer Lukas sieht sich gespannt in der geschmückten Turnhalle um.
2	Als erstes sieht er eine gruselige Hexe und gleich daneben unterhalten sich ein Cowboy und eine Prinzessin.	Zuerst sieht er eine gruselige Hexe und gleich daneben plaudern ein Cowboy und eine Prinzessin.	Als erstes sieht er eine gruselige Hexe und gleich daneben sind ein Cowboy und eine Prinzessin.	Als letztes sieht er eine gruselige Hexe und gleich daneben unterhalten sich ein Cowboy und eine Prinzessin.
3	Er bemerkt, dass es seine drei besten Freunde sind, als sie ihm zuwinken.	Er erkennt, dass es seine drei engsten Freunde sind, als sie ihm zuwinken.	Er bemerkt, dass er die drei bestens kennt, als sie ihm zuwinken.	Er bemerkt nicht, dass es seine drei besten Freunde sind, als sie ihm zuwinken.
4	Schnell pirscht Lukas sich an und bewirft die drei mit einer Ladung Konfetti.	Schnell pirscht Lukas sich an und beschmeißt die drei mit einer Ladung Konfetti.	Schnell pirscht Lukas sich an und ärgert die drei mit einer Ladung Konfetti.	Schnell pirscht Lukas sich an und bewirft die drei mit einer Ladung Bonbons.
5	Da zieht der Cowboy seine Wasserpistole und spritzt den verdutzten Lukas nass.	Da zieht der Cowboy seine Wasserpistole und macht den überraschten Lukas nass.	Da zieht der Cowboy seine Wasserpistole und schon ist der verdutzte Lukas nass.	Da zieht der Cowboy seine Wasserpistole und spritzt den verschmutzten Lukas nass.
6	Lukas ergreift die Flucht und so jagen die beiden durch die Menge.	Lukas ergreift die Flucht und so jagen die zwei durch die Menge.	Lukas ergreift die Flucht und so jagen die beiden durch die Turnhalle.	Lukas ergreift die Flucht und so jagen die vier durch die Menge.

Table D6

Original sentences, paraphrases, meaning changes, and situation changes of the story Die Renovierung ("The Renovation")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Weil Kathi und Annika ihr Zimmer streichen dürfen, fahren sie mit ihrem Vater Farbe kaufen.	Weil Kathi und Annika Zimmer anstreichen dürfen, fahren sie mit ihrem Vater Wandfarbe kaufen.	Weil Kathi und Annika ihr Zimmer verschönern dürfen, fahren sie mit ihrem Vater Farbe kaufen.	Weil Kathi und Annika ihr Gartenhaus streichen dürfen, fahren sie mit ihrem Vater Farbe kaufen.
2	Als sie ins Geschäft gehen, sagt der Vater, dass sie sich eine Farbe aussuchen sollen.	Als sie ins Geschäft gehen, sagt der Vater, dass sie …… eine Farbe auswählen sollen.	Als sie ins Geschäft gehen, erlaubt der Vater, dass sie sich eine Farbe aussuchen.	Als sie ins Geschäft gehen, sagt der Vater, welche Farbe sie aussuchen sollen.
3	Schon bald streiten die beiden heftig, weil Kathi lieber ein grünes und Annika lieber ein rotes Zimmer haben möchte.	Schon bald streiten die beiden heftig, weil Kathi lieber einen grünen und Annika lieber einen roten Raum haben möchte.	Schon bald streiten die beiden heftig, weil Kathi lieber grüne und Annika lieber rote Farbe haben möchte.	Schon bald streiten die beiden heftig, weil Kathi lieber ein grünes und Annika lieber ein rotes Gartenhaus haben möchte.
4	Bald wird der Vater wütend und droht ihnen, ohne Farbe zu gehen, wenn sie nicht gleich still sind.	Bald wird der Vater wütend und droht ihnen, ohne Farbe zu gehen, wenn sie nicht gleich ruhig sind.	Bald wird der Vater wütend und droht ihnen, ohne Farbe zu gehen, wenn sie nicht gleich aufhören.	Bald wird der Vater wütend und droht ihnen, ohne sie zu gehen, wenn sie nicht gleich still sind.
5	Plötzlich hat Kathi eine Idee und schnappt sich einfach zwei Farbeimer.	Auf einmal hat Kathi einen Gedanken und schnappt sich einfach zwei Farbeimer.	Plötzlich hat Kathi die Lösung und schnappt sich einfach zwei Farbeimer.	Plötzlich hat Kathi keine Lust mehr und schnappt sich einfach zwei Farbeimer.
6	Wieder zu Hause streichen sie eine Wand grün und eine andere rot und beide Mädchen sind zufrieden.	Wieder daheim streichen sie eine Wand grün und eine andere rot und beide Mädchen sind zufrieden.	Wieder zu Hause streichen sie eine Wand in Kathis und eine andere in Annikas Farbe und beide Mädchen sind zufrieden.	Wieder zu Hause streichen sie die Decke grün und eine andere Wand rot und beide Mädchen sind zufrieden.

Table D7

Original sentences, paraphrases, meaning changes, and situation changes of the story Anna und das Ballett ("Anna and the Ballet")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Kurz vor dem Auftritt der Ballettgruppe ist der Vorhang noch zu und Anna und ihre Freunde laufen hinter der Bühne wild durcheinander.	Kurz bevor die Ballettgruppe auftritt, ist der Vorhang noch zu und Anna und ihre Freunde laufen hinter der Bühne wild durcheinander.	Kurz vor dem Auftritt der Ballettgruppe ist der Vorhang noch zu und Anna und die anderen laufen hinter der Bühne wild durcheinander.	Kurz vor der Probe der Ballettgruppe ist der Vorhang noch zu und Anna und ihre Freunde laufen hinter der Bühne wild durcheinander.
2	Jeder geht noch einmal seine Rolle durch und alle sind sehr aufgeregt.	Jeder wiederholt noch einmal seine Rolle und alle sind sehr aufgeregt.	Jeder geht noch einmal seine Tanzschritte durch und alle sind sehr aufgeregt.	Jeder geht noch einmal seine Texte durch und alle sind sehr aufgeregt.
3	Anna rennt schnell noch einmal in den Flur und sieht ein letztes Mal in den Spiegel.	Anna rennt schnell noch einmal in den Gang und schaut ein letztes Mal in den Spiegel.	Anna rennt schnell noch einmal in den Flur und prüft ein letztes Mal ihr Spiegelbild.	Anna rennt schnell noch einmal in den Flur und sieht zum vorletzten Mal in den Spiegel.
4	Sie bemerkt, dass sich eine Haarsträhne gelöst hat und ärgert sich deswegen sehr.	Sie sieht, dass sich eine Haarsträhne gelockert hat und ärgert sich daher sehr.	Sie bemerkt, dass sich ihre Frisur gelöst hat und ärgert sich deswegen sehr.	Sie bemerkt, dass sich der Haarschmuck gelöst hat und ärgert sich deswegen sehr.
5	Sie versucht zu retten, was zu retten ist und nimmt schnell einen Kamm zur Hand, aber es ist zu spät.	Sie probiert zu retten, was zu retten ist und nimmt schnell einen Kamm zur Hand, aber es ist zu spät.	Sie versucht zu retten, was zu retten ist und nimmt schnell einen Kamm zur Hand, aber da muss sie los.	Sie versucht zu retten, was zu retten ist und nimmt schnell einen Kamm zur Hand, aber es ist zu schwer.
6	Sie hört ihre Musik und so tritt sie wohl oder übel hinaus ins Scheinwerferlicht.	Sie hört ihre Musik und so geht sie wohl oder übel hinaus ins Rampenlicht.	Sie hört ihre Musik und so tritt sie wohl oder übel hinaus auf die Bühne.	Sie hört ihre Musik und so tritt sie wohl oder übel in die Garderobe.

Table D8

Original sentences, paraphrases, meaning changes, and situation changes of the story Im Stadion ("In the Stadium")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Gespannt sitzen Basti und sein Vater auf der Tribüne, während sich die Fußballspieler auf dem Rasen warm machen.	Gespannt sitzen Basti und sein Vater auf der Tribüne, während sich die Fußballer auf dem Rasen aufwärmen.	Gespannt sitzen Basti und sein Vater auf der Tribüne, während sich die Fußballspieler auf dem Rasen warm laufen.	Gespannt sitzen Basti und sein Vater auf der Tribüne, während sich die Fußballspieler in der Kabine warm machen.
2	Fast alle Zuschauer tragen schwarz-gelbe Schals und auch Basti bindet sich den Schal seiner Lieblingsmannschaft um.	Fast jeder Zuschauer trägt einen schwarz-gelben Schal und auch Basti bindet sich den Schal seiner Lieblingsmannschaft um.	Fast alle Zuschauer haben schwarz-gelbe Schals besorgt und auch Basti bindet sich den Schal seiner Lieblingsmannschaft um.	Fast alle Spieler tragen schwarz-gelbe Schals und auch Basti bindet sich den Schal seiner Lieblingsmannschaft um.
3	Als er sich genauer umschaut, entdeckt er nur einige Fans der anderen Mannschaft.	Als er sich genauer umschaut, entdeckt er nur einige Anhänger der anderen Mannschaft.	Als er sich genauer umschaut, entdeckt er nur einige Fans in anderen Farben	Als er sich genauer umschaut, entdeckt er sehr viele Fans der anderen Mannschaft,
4	Vom lauten Anfeuern bekommt er einen ganz trockenen Hals und so reicht ihm der Vater einen Becher Saft.	Vom lauten Anfeuern bekommt er einen ganz trockenen Hals und so gibt der Vater ihm einen Becher Saft.	Vom lauten Anfeuern bekommt er einen ganz trockenen Hals und so reicht ihm der Vater etwas zu trinken.	Vom lauten Anfeuern bekommt er einen ganz trockenen Hals und so reicht ihm der Vater einen Becher Eis.
5	Auch der Vater hat schon einen ganz roten Kopf, weil heute so ein wichtiger Tag ist.	Auch der Vater hat schon einen ganz roten Kopf, weil der heutige Tag so wichtig ist.	Auch der Vater hat schon einen ganz roten Kopf, weil heute so ein wichtiges Spiel ist.	Auch der Vater hat schon einen ganz roten Kopf, weil heute so ein sonniger Tag ist.
6	Der Anpfiff ertönt und Basti drückt seiner Mannschaft ganz fest die Daumen.	Der Anpfiff erfolgt und Basti drückt seiner Mannschaft ganz fest die Daumen.	Das Spiel beginnt und Basti drückt seiner Mannschaft ganz fest die Daumen.	Der Gong ertönt und Basti drückt seiner Mannschaft ganz fest die Daumen.

Table D9

Original sentences, paraphrases, meaning changes, and situation changes of the story Reisevorbereitungen mit Hindernissen ("Travel Preparation with Obstacles")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Morgen ganz früh soll es für Familie Schmidt endlich mit dem vollgepackten Auto in den Urlaub gehen.	Am nächsten Tag ganz zeitig soll es für Familie Schmidt mit dem vollgepackten Auto endlich in den Urlaub gehen.	Morgen ganz früh soll es für Familie Schmidt endlich mit dem vollgepackten Auto auf die Reise gehen.	Übermorgen ganz früh soll es für Familie Schmidt endlich mit dem vollgepackten Auto in den Urlaub gehen.
2	Als Florian die letzten Koffer verlädt, entdeckt er überrascht, dass der rechte Hinterreifen des Autos ein Loch hat.	Als Florian die letzten Koffer verlädt, entdeckt er verblüfft, dass der rechte Hinterreifen des Autos ein Loch hat.	Als Florian die letzten Koffer verlädt, entdeckt er überrascht, dass der rechte Hinterreifen des Autos keine Luft hat.	Als Florian die letzten Koffer verlädt, entdeckt er überrascht, dass der rechte Hinterreifen des Autos weg ist.
3	Er ist ganz traurig, weil er sich schon so arg auf den Urlaub gefreut hat.	Er ist ganz traurig, weil er sich auf den Urlaub bereits so sehr gefreut hat.	Er ist ganz traurig, weil er sich schon so arg auf die Reise gefreut hat.	Er ist ganz traurig, weil er sich schon so arg auf die Freunde gefreut hat.
4	Schnell holt Florian seinen Vater, damit der den Reifen reparieren kann.	Schnell holt Florian seinen Vater, damit der den Reifen in Ordnung bringen kann.	Schnell holt Florian seinen Vater, damit der das Auto reparieren kann.	Schnell holt Florian seinen Vater, damit der einen Reifen kaufen kann.
5	Der Vater hebt mit seinem Wagenheber das Auto hoch und wechselt den Reifen.	Der Vater hebt mit seinem Wagenheber das Auto hoch und tauscht den Reifen aus.	Der Vater hebt mit seinem Wagenheber das Auto hoch und behebt den Schaden.	Der Vater hebt mit seinem Wagenheber das Auto hoch und wechselt die Bremsen.
6	Damit kann morgen der Urlaub losgehen und Florian ist sehr glücklich.	Damit kann der Urlaub morgen beginnen und Florian ist sehr glücklich.	Damit kann morgen die Familie losfahren und Florian ist sehr glücklich.	Damit kann morgen die Klassenfahrt losgehen und Florian ist sehr glücklich.

Table D10

Original sentences, paraphrases, meaning changes, and situation changes of the story Besuch für Sandra *(Visitors for Sandra)*

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Sandra öffnet ihrer Tante die Haustür und die beiden umarmen sich freudig.	Sandra öffnet ihrer Tante die Haustür und beide umarmen sich erfreut.	Sandra öffnet ihrer Tante die Haustür und die beiden begrüßen sich freudig.	Sandra öffnet ihrer Tante die Haustür und die beiden kichern freudig.
2	Die Tante trägt einen geheimnisvollen Korb unterm Arm und Sandra versucht hineinzuschauen.	Ihre Tante hat einen geheimnisvollen Korb unterm Arm und Sandra versucht hineinzuschauen.	Die Tante bringt einen geheimnisvollen Korb mit und Sandra versucht hinzuschauen.	Die Mutter trägt einen geheimnisvollen Korb unterm Arm und Sandra versucht hineinzuschauen.
3	Auf einmal bewegt sich der Korb und Sandra springt erschrocken zurück.	Auf einmal wackelt der Korb und Sandra springt erschrocken zurück.	Auf einmal bewegt sich etwas in dem Korb und Sandra springt erschrocken zurück.	Auf einmal bewegt sich die Tante und Sandra springt erschrocken zurück.
4	Da hört sie ein Winseln aus dem Korb und ein kleiner Hund streckt seinen Kopf heraus.	Da vernimmt sie ein Winseln aus dem Korb und ein kleiner Hund streckt seinen Kopf heraus.	Da hört sie Laute aus dem Korb und ein kleiner Hund streckt seinen Kopf heraus.	Da hört sie ein Bellen aus dem Korb und ein kleiner Hund streckt seinen Kopf heraus.
5	Sandra streckt ihm vorsichtig ihre Hand entgegen und der kleine Hund schleckt die Hand zutraulich ab.	Sandra hält ihm behutsam ihre Hand hin und der kleine Hund schleckt die Hand zutraulich ab.	Sandra streckt ihm langsam ihre Hand entgegen und der kleine Hund schleckt die Hand zutraulich ab.	Sandra streckt ihm stürmisch ihre Hand entgegen und der kleine Hund schleckt die Hand zutraulich ab.
6	Sie hebt ihn aus dem Korb und Sandra tollt gemeinsam mit dem Hund im Garten herum.	Sie nimmt den Hund aus dem Korb und Sandra tollt gemeinsam mit dem Hund im Garten herum.	Sie hilft ihm aus dem Korb und Sandra tollt gemeinsam mit dem Hund im Garten herum.	Ihre Tante hebt ihn aus dem Korb und Sandra tollt gemeinsam mit dem Hund im Garten herum.

Table D11

Original sentences, paraphrases, meaning changes, and situation changes of the story Maries Geburtstag ("Marie's Birthday")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Schlaflos wälzt sich Marie in ihrem Bett, denn morgen wird sie zehn Jahre alt.	Schlaflos wälzt sich Marie in ihrem Bett, denn morgen wird sie zehn.	Schlaflos wälzt sich Marie in ihrem Bett, denn morgen wird sie ein Jahr älter.	Schlaflos wälzt sich Marie in ihrem Bett, denn gestern wurde sie zehn Jahre alt.
2	Sie denkt an die vielen Geschenke und die lieben Freunde, die zu Kakao und Kuchen kommen.	Sie denkt über die zahlreichen Geschenke und die netten Freunde nach, die zu Kakao und Kuchen kommen.	Sie denkt an die vielen Geschenke und die lieben Gäste, die zu Kakao und Kuchen kommen.	Sie denkt an die vielen Spiele und die lieben Freunde, die zu Kakao und Kuchen kommen.
3	Um sechs hält sie es in ihrem Bett nicht mehr aus und sie steht auf.	Um sechs erträgt sie es nicht mehr in ihrem Bett und sie steht auf.	Früh morgens hält sie es in ihrem Bett nicht mehr aus und sie steht auf.	Um zehn hält sie es in ihrem Bett nicht mehr aus und sie steht auf.
4	Sie zieht gerade ihr Lieblingskleid aus dem Schrank, als sie auf dem Flur leise Stimmen hört.	Sie zieht gerade ihr Lieblingskleid aus dem Schrank, als sie leise Stimmen auf dem Gang hört.	Sie zieht gerade ihr Lieblingskleid aus dem Schrank, als sie auf dem Flur leise Geräusche hört.	Sie zieht gerade ihr Lieblingskleid aus dem Schrank, als sie vor dem Haus leise Stimmen hört.
5	Da öffnet sich ihre Zimmertür und ihre Eltern kommen mit einer sehr großen Torte herein.	Da öffnet sich ihre Zimmertür und ihre Eltern kommen mit einer riesigen Torte herein.	Da öffnet sich ihre Zimmertür und ihre Eltern überraschen sie mit einer sehr großen Torte.	Da öffnet sich ihre Zimmertür und ihre Eltern kommen mit einem sehr großen Geschenk herein.
6	Marie bläst alle Kerzen auf Anhieb aus und wünscht sich ganz fest, dass sie jeden Tag Geburtstag hat.	Marie bläst alle Kerzen auf Anhieb aus und wünscht sich ganz stark, dass jeder Tag ihr Geburtstag ist.	Marie bläst alle Kerzen auf Anhieb aus und wünscht sich ganz fest, dass sie jeden Tag Geschenke bekommt.	Marie bläst alle Kerzen auf Anhieb aus und wünscht sich ganz fest, dass sie jeden Tag schulfrei hat.

Table D12

Original sentences, paraphrases, meaning changes, and situation changes of the story Simon kommt zu spät ("Simon's Too Late")

no.	Original	Paraphrase	Meaning Change	Situation Change
1	Heute wird ein Klassenfoto gemacht und so hüpft Simon die Stufen hinunter und rennt auf den Schulhof.	Heute wird die Klasse fotografiert und so hüpft Simon die Stufen hinunter und rennt auf den Schulhof.	Heute wird ein Gruppenfoto gemacht und so hüpft Simon die Stufen hinunter und rennt auf den Schulhof.	Heute wird ein Klassenfoto gemacht und so hüpft Simon die Stufen hinunter und rennt in die Turnhalle.
2	Da steht nämlich schon seine ganze Klasse bereit und wartet nur auf ihn.	Da steht nämlich schon seine ganze Klasse bereit und wartet bloß noch auf ihn.	Da steht nämlich schon seine ganze Klasse bereit und wartet ungeduldig auf ihn.	Da steht nämlich schon seine ganze Klasse bereit und wartet auf den Fotografen.
3	Simon entschuldigt sich bei der Lehrerin und stellt sich schnell neben Martin in die erste Reihe.	Simon entschuldigt sich bei der Lehrerin und stellt sich schnell an Martins Seite in die erste Reihe.	Simon entschuldigt sich bei der Lehrerin und stellt sich schnell neben Martin nach vorne.	Simon entschuldigt sich bei der Lehrerin und stellt sich schnell vor Martin in die erste Reihe.
4	Doch da beschwert sich die Nervensäge Sarah, weil sie hinter Simon gar nicht zu sehen ist.	Doch da beschwert sich die Nervensäge Sarah, weil man sie hinter Simon gar nicht sehen kann.	Doch da beschwert sich die Nervensäge Sarah, weil Simon zu groß ist.	Doch da beschwert sich die Nervensäge Sarah, weil sie hinter Simon gar nicht zu hören ist.
5	Wie immer muss Simon also in die hinterste Reihe und ist deshalb sehr sauer.	Wie jedes Mal muss Simon also in die letzte Reihe und ist deshalb sehr sauer.	Wie schon letztes Mal muss Simon also in die hinterste Reihe und ist deshalb sehr sauer.	Zum ersten Mal muss Simon also in die hinterste Reihe und ist deshalb sehr sauer.
6	Er ist darum auch der einzige, der nicht lächelt, als der Fotograf dann das Foto macht.	Er ist darum auch die einzige Person, die nicht lächelt, als der Fotograf dann das Foto macht.	Er ist darum auch der einzige, der nicht freundlich schaut, als der Fotograf dann das Foto macht.	Er ist darum auch der einzige, der nicht in die Kamera schaut, als der Fotograf dann das Foto macht.

Appendix D3: Instruction and practice trial

Original instruction	English translation
Schön, dass du heute hier bist und mitmachst! Du bekommst zwei Aufgaben. Bei der ersten wirst du kurze Geschichten lesen (*Experiment 1*)/hören (*Experiment 2*), die du dir genau merken sollst. Denn nach vier Geschichten bist du an der Reihe! Da bekommst du einzelne Sätze und sollst entscheiden, ob die Sätze in einer der Geschichten vorgekommen sind. Wenn ja, dann drückst du das Häkchen rechts. Wenn du glaubst, der Satz kam nicht in der Geschichte vor, drückst du das Kreuz hier links. Insgesamt sind es 12 Geschichten, und nach vier Geschichten, wenn du die Aufgabe gemacht hast, kannst du eine kurze Pause machen. Das Ganze dauert etwa eine halbe Stunde.	Nice to have you here taking part! You'll get two tasks. In the first one, you're going to read (*Experiment 1*)/listen to (*Experiment 2*) short stories that you should remember exactly. Because after four stories, it's your turn! Then you're going to get single sentences and you'll have to decide if these sentences have happened in one of the stories. If so, then you press the check mark to the right. If you think the sentence didn't happen in the story, you press the cross here to the left. In total, there are 12 stories, and after every four stories, when you've performed the task, you can take a short break. The whole thing's going to take about half an hour.
Als erstes machen wir jetzt eine Übung zusammen, bevor es richtig losgeht. Hast du noch Fragen?	So let's start first with a practice trial before we get going. Have you got any questions?
Gegebenenfalls Fragen klären	*Resolve questions, if any*
Dann setze bitte jetzt die Kopfhörer auf. *(nur Experiment 2)*	So please take now your headphones on. *(Experiment 2 only)*
*Experiment 1: Versuchsleiter*in stellt Übungsgeschichte vor*	*Experiment 1: Experimenter introduces practice story*
Dann liest du dir jetzt bitte erst mal die Übungsgeschichte durch, Satz für Satz. Jedes Mal, wenn du mit dem nächsten Satz weiterlesen möchtest, drücke die Taste mit dem Buch-Symbol.	So now, please start with reading the practice story, sentence per sentence. Every time you want to go on reading the next sentence, push the key with the book symbol.
Experiment 2: Aufgenommene Sprecherin stellt sich und die Übungsgeschichte vor	*Experiment 2: Recorded speaker introduces herself and the practice story*
Hallo, ich heiße Katharina und freue mich, dass du heute mit dabei bist. Dir wurde ja schon erklärt, was du heute machen sollst, deshalb können wir auch gleich loslegen. Zuerst lese ich dir eine Übungsgeschichte vor. Also pass gut auf!	Hello, my name is Katharina and I'm happy to have you here today. You have been told what you are going to do today, that's why we can start right now. At first, I'm going to read you a practice story. So listen carefully!

Übungsgeschichte	*Practice story*
Annemarie freut sich, denn ihre große Schwester spielt heute mit ihr.	Annemarie is happy because her big sister's playing with her today.
Zuerst verkleiden sie sich mit den Kleidern ihrer Mutter und schminken sich vor dem großen Spiegel.	First, they dress up with their mother's clothes and paint their faces in front of the big mirror.
Danach gehen sie raus in den Garten und schaukeln ganz wild auf der Schaukel.	Then, they go outside in the garden and see-saw wildly on the swings.
So, das war auch schon die Geschichte von Annemarie. Jetzt bist du mit der Fehlersuche dran. Los geht's.	So that was the story about Annemarie. Now, it's your turn to start the fault finding. Here you go!
Übungsaufgabe 1 (situation change)	*Practice trial 1 (situation change)*
Annemarie freut sich, denn ihre Cousine spielt heute mit ihr.	Annemarie is happy because her cousin's playing with her today.
Übungsaufgabe 2 (original sentence)	*Practice trial 1 (original sentence)*
Danach gehen sie raus in den Garten und schaukeln ganz wild auf der Schaukel.	Then, they go outside in the garden and see-saw wildly on the swings.
Übungsaufgabe 3 (paraphrase)	*Practice trial 3 (paraphrase)*
Zuerst verkleiden sie sich mit den Klamotten ihrer Mutter und schminken sich vor dem großen Spiegel.	First, they dress up with their mother's gear and paint their faces in front of the big mirror.
*Experiment 1: Versuchsleiter*in*	*Experiment 1: Experimenter*
Okay. Dann geht's jetzt richtig los! Aber Achtung: Die neuen Geschichten sind ein bisschen länger als die, die du dir gerade durchgelesen hast.	Okay. So now let's get going! But be careful: The new stories are a bit longer than the one you just have read.
Experiment 2: Aufgenommene Sprecherin	*Experiment 2: Recorded speaker*
Okay. Dann geht's jetzt richtig los! Aber Achtung: Die neuen Geschichten sind ein bisschen länger als die, die ich dir gerade vorgelesen hab.	Okay. So now let's get going! But be careful: The new stories are a bit longer than the one I just have read to you.
*Es wird kein Feedback zur Übungsaufgabe gegeben. Sofern der*die Versuchsleiter*in es für nötig erachtet, werden Instruktionen wiederholt oder genauer erläutert.*	*No feedback is given with regard to the practice trial. If the experimenter deems it necessary, instructions are repeated or explained.*

Appendix E: Material of Experiment 3

Appendix E1: Stories, illustrations, and target pictures

Appendix E1.1: Target trials with downward movement

no.	Sentence	Picture	no.	Sentence	Picture
1	An einem schönen Spätsommertag geht Daniel hinaus in den Garten und setzt sich unter einen Baum.		4	Er steht auf und beginnt, den Baum kräftig zu schütteln.	
2	Er blickt nach oben und sieht einen saftigen Apfel an einem der hohen Zweige hängen.		5	Der Apfel fällt hinunter auf das Gras.	
3	Daniel kann nicht hinaufklettern, doch da kommt ihm eine Idee.			*Target (positive)*	

Figure E1. Target story "Apfel" (apple).

no.	Sentence	Picture	no.	Sentence	Picture
1	Laura möchte eine Fahrradtour machen und steht schon abfahrbereit vor dem Haus.		4	Laura steigt auf einen Stuhl neben dem Regal und versucht, hinaufzugreifen.	
2	Ihre kleine Schwester kommt herausgerannt und sagt Laura, dass sie ihren Fahrradhelm vergessen hat.		5	Dabei fällt der Fahrradhelm auf den Boden hinab.	
3	Ob er wohl noch oben auf dem Regal liegt?			*Target (positive)*	

Figure E2. Target story "Fahrradhelm" (bike helmet).

no.	Sentence	Picture	no.	Sentence	Picture
1	Markus und sein Vater gehen im Wald wandern.		4	Plötzlich kommt Wind auf und fegt durch die Tannen.	
2	Nach einiger Zeit werden sie müde und beschließen, auf einer schattigen Bank Rast zu machen.		5	Da fällt ein Zapfen neben Markus hinunter auf den Waldboden.	
3	Die beiden setzen sich und kramen ihr Pausenbrot aus den Rucksäcken hervor.			*Target (positive)*	

Figure E3. Target story "Zapfen" (cone).

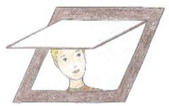
no.	Sentence	Picture	no.	Sentence	Picture
1	Weil es heute sehr heiß war, öffnet Robert das Dachfenster in seinem Zimmer, um zu lüften.		4	Die Wolken kommen immer näher, und so macht Robert schnell das Dachfenster zu.	
2	Er steht am Fenster und genießt die kühle Abendluft in seinem Gesicht.		5	Schon bald prasseln die ersten Regentropfen auf die Fensterscheibe.	
3	In der Ferne sieht er dunkle Wolken und hört den Donner grollen.			*Target (positive)*	

Figure E4. Target story "Regentropfen" (raindrops).

no.	Sentence	Picture	no.	Sentence	Picture
1	Sophie sitzt auf dem Sofa und liest ein spannendes Buch.		4	Als sie näherkommt, erkennt sie eine Spinne in ihrem Netz.	
2	Als sie gerade eine kurze Pause macht, schaut sie in die Ecke.		5	Auf einmal seilt sich die Spinne von der Decke ab.	
3	Dort entdeckt sie etwas, und so steht sie auf, um genauer hinzusehen.			*Target (positive)*	

Figure E5. Target story “Spinne” (spider).

no.	Sentence	Picture	no.	Sentence	Picture
1	Es ist Heiligabend, und Tobias und seine Familie schmücken den Weihnachtsbaum.		4	Schnell läuft Tobias zum Ofen, um im Schornstein nachzusehen.	
2	Auf einmal hört Tobias ein Klappern, das immer näher kommt.		5	Da rutscht der Weihnachtsmann auch schon durch den Schornstein hinunter.	
3	Ob das wohl der Weihnachtsmann mit seinem Rentierschlitten ist?			*Target (positive)*	

Figure E6. Target story “Weihnachtsmann” (Santa Claus).

no.	Sentence	Picture	no.	Sentence	Picture
1	Schlecht gelaunt sitzt Fabian an seinem Schreibtisch, weil er seine Hausaufgaben noch nicht fertig hat.		4	Doch er versteht die Aufgaben nicht und wird wütend.	
2	Er hat sich mit seinen Freunden zum Fußball-spielen verabredet, und alle warten schon auf ihn.		5	Er steht auf und wirft das Buch hinunter auf den Boden.	
3	Als letztes ist Mathe dran, und so holt er das Mathebuch aus seiner Tasche.			*Target (positive)*	

Figure E7. Target story "Buch" (book).

no.	Sentence	Picture	no.	Sentence	Picture
1	Als früh morgens der Wecker klingelt, ist Marie noch sehr müde.		4	Mürrisch steht sie auf und streckt sich.	
2	Gerade eben hat sie noch von dem schönen Familienausflug geträumt.		5	Dann wirft sie voller Wut den Wecker hinunter auf den Teppichboden.	
3	Doch heute ist Montag und Marie muss wieder in die Schule.			*Target (positive)*	

Figure E8. Target story "Wecker" (alarm clock).

no.	Sentence	Picture	no.	Sentence	Picture
1	Käpt'n Tom und seine Piratenmannschaft haben auf einer Insel eine Schatztruhe entdeckt.		4	Käpt'n Tom muss schnell etwas tun, denn sonst gehen sie alle unter!	
2	Während die Mannschaft lossegelt, zählt er die erbeuteten Goldstücke.		5	Und so lässt er die Schatztruhe ins Meer hinabsinken.	
3	Doch bald schon merken die Piraten, dass ihr Schiff zu schwer geworden ist.			*Target (positive)*	

Figure E9. Target story "Schatztruhe" (treasure chest).

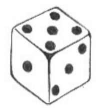
no.	Sentence	Picture	no.	Sentence	Picture
1	Voller Vorfreude öffnet Leonie die Spieleschachtel, die sie von ihrer Tante geschenkt bekommen hat.		4	Leonie schüttelt kräftig und wartet gespannt darauf, was sie bekommen wird.	
2	Erstaunt erblickt sie die Figuren und Kärtchen und all die anderen wunderbaren Dinge in der Schachtel.		5	Mit Schwung haut sie den Würfelbecher hinab auf den Tisch.	
3	Als nächstes sieht sie einen Würfel und legt ihn in einen Würfelbecher.			*Target (positive)*	

Figure E10. Target story "Würfel" (dice).

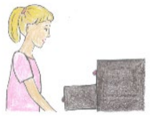
no.	Sentence	Picture	no.	Sentence	Picture
1	Gleicht geht Annika zum Basteln, aber sie findet ihren Klebestift nicht.		4	Verzweifelt wühlt Annika in ihrem Bastelkorb herum.	
2	Sie fragt ihre Mutter, ob sie ihn gesehen hat.		5	Dabei lässt sie eine Schere auf den Boden fallen.	
3	Die Mutter antwortet genervt, dass sie einfach mal in ihrem Bastelkorb nachsehen soll.			*Target (positive)*	

Figure E11. Target story “Schere” (scissors).

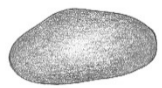
no.	Sentence	Picture	no.	Sentence	Picture
1	Heute ist ein heißer Tag, und so setzt sich Ingo auf die Wiese und lehnt sich gegen den Brunnen.		4	Als er das Geräusch wieder hört, steht Ingo be-unruhigt auf und nimmt einen großen Stein.	
2	Plötzlich hört er ein seltsames, dumpfes Ge-räusch.		5	Er wirft den Stein hinab in den Brunnen.	
3	Ob da wohl etwas in dem Brunnen ist?			*Target (positive)*	

Figure E12. Target story “Stein” (stone).

Appendix E1.2: Target trials with upward movement

no.	Sentence	Picture	no.	Sentence	Picture
1	Heute hat der Herbst begonnen und die Bäume sind schon bunt gefärbt.		4	Schon bald kommt ein frischer Wind und wirbelt die Blätter durch die Luft.	
2	Jonas und Corinna ziehen ihre Mäntel an und gehen mit ihrem Drachen in den Park.		5	Jetzt steigt der Drachen hinauf in den Himmel.	
3	Die beiden Geschwister steigen auf einen Hügel und Jonas hält die Schnüre fest.			*Target (positive)*	

Figure E13. Target story “Drachen” (kite).

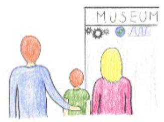
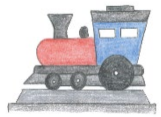
no.	Sentence	Picture	no.	Sentence	Picture
1	Timo und seine Eltern gehen ins Technikmuseum und schauen sich ganz viele Maschinen an.		4	Timo ist inzwischen weitergelaufen zu einer großen Leinwand, wo ein Film über Raumfahrt zu sehen ist.	
2	Gleich am Eingang steht eine große, alte Dampflok.		5	In dem Film startet gerade eine Rakete hoch ins Weltall.	
3	Voller Begeisterung bleiben die Eltern vor der Lok stehen und betrachten sie ganz lange.			*Target (positive)*	

Figure E14. Target story “Rakete” (rocket).

no.	Sentence	Picture	no.	Sentence	Picture
1	Leon und sein großer Bruder Andy sind heute auf dem Flugplatz, weil sie eine Fahrt im Hubschrauber geschenkt bekommen haben.		4	Als der Pilot den Motor anwirft, schaut Leon seinem Bruder vom Flugplatz aus zu.	
2	Andy freut sich total auf den Flug, aber Leon traut sich nicht, mitzufliegen.		5	Der Hubschrauber hebt ab und steigt in den Himmel auf.	
3	Vergeblich versucht Andy, seinem Bruder Mut zu machen und steigt danach alleine ein.			*Target (positive)*	

Figure E15. Target story “Hubschrauber” (helicopter).

no.	Sentence	Picture	no.	Sentence	Picture
1	An einem freundlichen Nachmittag gehen Vera und ihr Vater auf einem Feldweg spazieren.		4	Vera macht große Augen - wie gern würde sie mitfliegen!	
2	Auf ihrem Weg sehen sie einen Heißluftballon am Boden stehen.		5	Plötzlich steigt der Heißluftballon hoch in die Wolken.	
3	Sie winken den Leuten zu, die in dem Ballon sitzen.			*Target (positive)*	

Figure E16. Target story “Heißluftballon” (hot-air balloon).

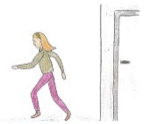
no.	Sentence	Picture	no.	Sentence	Picture
1	Am Silvesterabend sitzen die Eltern noch am Tisch, während Carla bereits ihre Schnürsenkel zubindet.		4	Die Glocken ertönen, und so rennt Carla hinaus auf die Straße.	
2	Sie hat Angst, das Feuerwerk zu verpassen und öffnet schnell die Wohnungstür.		5	Da fliegt auch schon die ersten Feuerwerksrakete pfeifend in den Himmel.	
3	Auch aus den anderen Häusern kommen schon Leute und stellen sich dicht an dicht auf die Straße.			*Target (positive)*	

Figure E17. Target story "Feuerwerksrakete" (firework rocket).

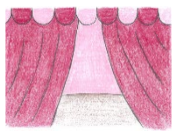
no.	Sentence	Picture	no.	Sentence	Picture
1	Während Niklas und seine Mutter in letzter Minute Platz nehmen, beginnen die Zuschauer im Zirkus bereits zu klatschen.		4	Da kommt der Clown auf seinem Einrad heraus, mit einem Luftballon in der Hand.	
2	Niklas schaut erwartungsvoll in die Manege, während die Musikkapelle noch spielt.		5	Der Clown lässt die Schnur los und der Luftballon fliegt hinauf zur Zirkusdecke.	
3	Plötzlich hört die Musik auf und der Vorhang öffnet sich langsam.			*Target (positive)*	

Figure E18. Target story "Luftballon" (balloon).

no.	Sentence	Picture	no.	Sentence	Picture
1	Oliver und seine Mutter machen heute Frühstück.		4	Als er gerade die Butter aus dem Kühlschrank nehmen will, hört er hinter sich ein Klacken.	
2	Während die Mutter ihren Kaffee kocht, deckt Oliver den Tisch.		5	Oliver dreht sich um und zieht eine Toastscheibe schwungvoll aus dem Toaster.	
3	Eifrig holt er Teller und Besteck aus dem Schrank und breitet sie auf dem Esstisch aus.			*Target (positive)*	

Figure E19. Target story "Toastscheibe" (toast slice).

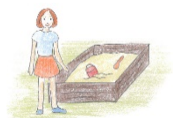
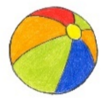
no.	Sentence	Picture	no.	Sentence	Picture
1	Claudia und ihre Freundin Lea toben auf dem Spielplatz herum.		4	Lea ist inzwischen einen Turm hochgeklettert und winkt ihrer Freundin zu.	
2	Lea hat einen Ball mitgebracht, den die beiden Mädchen sich immer wieder zuwerfen.		5	Claudia wirft den Ball zu Lea in die Höhe.	
3	Als nächstes ist Claudia an der Reihe, die gerade am Sandkasten steht.			*Target (positive)*	

Figure E20. Target story "Ball" (ball).

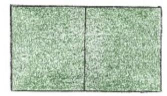
no.	Sentence	Picture	no.	Sentence	Picture
1	Gespannt sitzt die Klasse im Klassenzimmer und wartet auf die neue Lehrerin.		4	Frau Bauer begrüßt freundlich ihre neue Klasse und legt ihre Tasche auf das Lehrerpult.	
2	Sie heißt Frau Bauer und soll ziemlich streng sein.		5	Dann schiebt sie die Tafel nach oben.	
3	Da öffnet sich auch schon die Tür und sie kommt herein.			*Target (positive)*	

Figure E21. Target story “Tafel” (blackboard).

no.	Sentence	Picture	no.	Sentence	Picture
1	Familie Schuster macht Urlaub an der Nordsee, und Martin genießt das Baden im Meer.		4	Er bückt sich und sieht eine große Muschel im Sand liegen.	
2	Gerade kommt er heraus und macht sich erfrischt auf den Weg zum Strandkorb.		5	Martin hebt die Muschel auf.	
3	Doch auf dem Weg stößt er mit seinem Fuß an einen harten, spitzen Gegenstand.			*Target (positive)*	

Figure E22. Target story “Muschel” (shell).

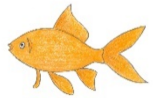
no.	Sentence	Picture
1	Heute geht Hannah mit ihren Eltern zum Angeln an den See.	
2	Doch bisher hat kein Fisch angebissen, und Papa hat außer viel Müll noch nichts geangelt.	
3	Am späten Nachmittag hat Papa keine Lust mehr und gibt Hannah die Angel.	
4	Kurze Zeit später macht die Angel einen Ruck, und Hannah wäre beinahe in den See gefallen.	
5	Da zieht sie einen großen Fisch aus dem Wasser.	
	Target (positive)	

Figure E23. Target story "Fisch" (fish).

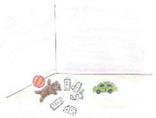
no.	Sentence	Picture
1	Morgen geht es in den Urlaub, und deshalb räumt Vanessa heute fleißig ihr Zimmer auf.	
2	Jetzt ist die hinterste Ecke dran, wo noch ganz viele Spielsachen kreuz und quer auf dem Boden herumliegen.	
3	Plötzlich entdeckt sie eine Puppe zwischen den Dominosteinen, die sie traurig anschaut.	
4	Da bekommt Vanessa Mitleid und räumt die Dominosteine in eine Schachtel.	
5	Danach legt sie die Puppe auf ihr Hochbett hinauf.	
	Target (positive)	

Figure E24. Target story "Puppe" (doll).

Appendix E1.3: Filler trials

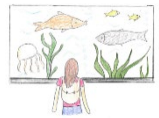
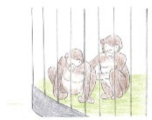
no.	Sentence	Picture	no.	Sentence	Picture
1	Heute geht die Klasse in den Zoo, und Judith freut sich am meisten auf die Meeresschild-kröten.		4	Schon bald kommt sie zum Aquarium, wo gerade eine Meeresschildkröte elegant durchs Wasser gleitet.	
2	Doch zuerst gehen sie ins Affenhaus, und die Kinder schauen begeistert zu, wie zwei Schimpansen sich lausen.		5	Zufrieden schaut Judith zu und hofft, dass die anderen ihre Flucht nicht bemerken.	
3	Als es Judith zu lange dauert, beschließt sie, einfach alleine weiterzugehen.			*Target (positive)*	

Figure E25. Filler story "Meeresschildkröte" (sea turtle).

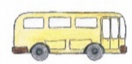
no.	Sentence	Picture	no.	Sentence	Picture
1	Heute ist ein ganz normaler Schultag, doch für Dominik fängt er gar nicht gut an.		4	Genau in dem Moment kommt der Schulbus um die Ecke gefahren, und Dominik flitzt los.	
2	Er muss sich nämlich sehr beeilen, damit er seinen Schulbus noch bekommt.		5	Doch als er gerade einsteigen will, stellt er fest, dass er seine Fahrkarte vergessen hat.	
3	Schnell packt er sein Pausenbrot ein und zieht seine Jacke an.			*Target (positive)*	

Figure E26. Filler story "Schulbus" (school bus).

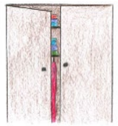
no.	Sentence	Picture	no.	Sentence	Picture
1	Am Morgen steht Lukas mit einem karierten Hemd, einer Weste und einer Spielzeugpistole in der Hosentasche vor dem Spiegel.		4	Da fällt ihm ein, dass in Mamas Kleiderschrank noch ein alter Hut liegt.	
2	Heute wird im Kindergarten nämlich Fasching gefeiert und Lukas möchte als Cowboy gehen.		5	Deshalb geht Lukas zu seiner Mutter und fragt sie, ob er ihn haben kann.	
3	Doch plötzlich stellt er fest, dass er gar keinen Hut hat!			*Target (positive)*	

Figure E27. Filler story “Hut” (hat).

no.	Sentence	Picture	no.	Sentence	Picture
1	Pia und Felix machen mit ihren Eltern Urlaub auf dem Campingplatz, und heute Abend gibt es ein Lagerfeuer.		4	Er läuft in den Wohnwagen und kommt kurze Zeit später mit seiner Gitarre heraus.	
2	Pia ist gut gelaunt, deshalb holt sie ihr Liederbuch aus der Tasche und fängt an zu singen.		5	Und so singt und musiziert die Familie am Lagerfeuer bis spät in die Nacht hinein.	
3	Da kommt Felix eine Idee und er steht auf.			*Target (positive)*	

Figure E28. Filler story “Gitarre” (guitar).

no.	Sentence	Picture	no.	Sentence	Picture
1	Simon ist mit seiner Mutter in der Stadt und möchte ein Eis haben.		4	Schon bald ist er auch an der Reihe und sucht sich eine große Kugel Vanilleeis aus.	
2	Doch die Mutter sagt nein, weil er heute Morgen zu spät in die Schule kam.		5	Er schaut zu seiner Mutter und verspricht ihr, morgen pünktlich zur Schule zu gehen.	
3	Simon nervt seine Mutter nun so lange, bis sie schließlich nachgibt und er sich bei der nächsten Eisdiele anstellen darf.			*Target (positive)*	

Figure E29. Filler story “Eis” (ice cream).

no.	Sentence	Picture	no.	Sentence	Picture
1	Eva und Sabrina sitzen mit ihrer Oma am Tisch und unterhalten sich.		4	Als Oma wieder am Tisch ist, nimmt Eva das Kartenspiel und mischt es sorgfältig.	
2	Plötzlich steht Oma auf, geht zum Spieleschrank und holt ein Kartenspiel heraus.		5	Ob sie wohl diesmal gewinnen wird?	
3	Da freuen sich die beiden Schwestern, denn gleich wird Mau-Mau gespielt.			*Target (positive)*	

Figure E30. Filler story “Kartenspiel” (card game).

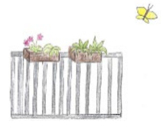
no.	Sentence	Picture	no.	Sentence	Picture
1	Jana und ihre Eltern sitzen beim Abendessen auf dem Balkon.		4	Als die Katze auf das Geländer klettert, ist der Schmetterling bereits weggeflogen.	
2	Die Katze sitzt auf Janas Schoß und schaut neugierig zum Balkongeländer hinüber.		5	Da springt die Katze hinunter auf den Gehweg.	
3	Sie beobachtet einen Schmetterling, der zwischen den Pflanzen in den Kästen herumfliegt.			*Target (negative)*	

Figure E31. Filler story "Katze" (cat).

no.	Sentence	Picture	no.	Sentence	Picture
1	Spät abends, als es schon dunkel ist, steht Anita vor der Haustür und sucht ihren Schlüssel.		4	Deshalb versucht sie auch sehr leise zu sein, während sie ihre Handtasche durchwühlt.	
2	Sie war bei ihrer besten Freundin und die beiden haben sehr viel miteinander geredet.		5	Plötzlich fällt der Schlüssel mit lautem Klimpern auf den Pflasterstein.	
3	Bestimmt sind die Eltern sauer, weil sie so lange weg war.			*Target (negative)*	

Figure E32. Filler story "Schlüssel" (key).

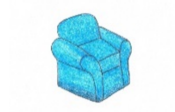
no.	Sentence	Picture	no.	Sentence	Picture
1	Peter sitzt auf der Wiese, als er plötzlich einen Maulwurf neben sich aus der Erde auftauchen sieht.		4	Peter und seine Mutter rennen schnell zu dem Erdhügel, wo der Maulwurf zu sehen war.	
2	Überrascht rennt er zum Haus, um seine Mutter zu holen.		5	Doch als sie gerade das Foto machen will, verschwindet er wieder unter der Erde.	
3	Kurze Zeit später kommt sie herausgeeilt, mit einer Kamera in der Hand.			*Target (negative)*	

Figure E33. Filler story “Maulwurf” (mole).

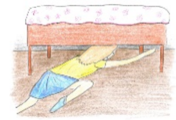
no.	Sentence	Picture	no.	Sentence	Picture
1	Es ist schon spät, doch Anna und Tamara sind überhaupt noch nicht müde.		4	Lachend springt Anna auf und versteckt sich unter ihrem Bett.	
2	Während Anna auf ihrem Bett mit den Kuscheltieren spielt, tobt Tamara im Kinderzimmer herum.		5	Tamara wirft das Kissen hinunter zu Anna.	
3	Plötzlich schnappt sich Tamara ein Kissen und rennt damit kreischend auf Anna zu.			*Target (negative)*	

Figure E34. Filler story “Kissen” (pillow).

no.	Sentence	Picture	no.	Sentence	Picture
1	Heute darf Jan zum ersten Mal beim Kochen helfen, und er ist schon ganz aufgeregt.		4	Jan nimmt das Ei aus der Schachtel und schlägt es vorsichtig am Pfannenrand auf.	
2	Es gibt Pellkartoffeln und Spinat, und Jan soll ein Spiegelei dazu braten.		5	Danach lässt er es in die Pfanne hineingleiten.	
3	Er holt die Pfanne aus dem Schrank und gießt etwas Öl hinein.			*Target (negative)*	

Figure E35. Filler story “Ei” (egg).

no.	Sentence	Picture	no.	Sentence	Picture
1	Stefan und sein Großvater sitzen am Früh-stückstisch und plaudern.		4	In dem Moment klingelt es an der Haustür.	
2	Vergnügt berichtet der Großvater aus seiner Schulzeit, während Stefan an seiner Teetasse nippt.		5	Großvater steht auf und lässt die Zeitung in den Zeitungshalter hinabfallen.	
3	Als der Großvater fertig ist mit Erzählen, kramt er seine Zeitung hervor und beginnt zu lesen.			*Target (negative)*	

Figure E36. Filler story “Zeitung” (newspaper).

no.	Sentence	Picture	no.	Sentence	Picture
1	An einem sonnigen Herbstmorgen hilft Max seinen Eltern bei der Gartenarbeit.		4	Der verdutzte Max versucht gerade, loszurennen und die Nuss zurückzubekommen.	
2	Vater und Mutter graben gerade das Gemüsebeet um, und Max darf Haselnüsse ernten.		5	Doch da klettert das Eichhörnchen bereits einen Baum hinauf.	
3	Während er die reifen Nüsse vom Boden aufsammelt, huscht plötzlich ein Eichhörnchen herbei und schnappt sich ein paar Nüsse aus dem Korb.			*Target (negative)*	

Figure E37. Filler story “Eichhörnchen” (squirrel).

no.	Sentence	Picture	no.	Sentence	Picture
1	Sandra und ihre Familie sind heute auf dem Flughafen, um sich von Tante Elisabeth zu verabschieden.		4	Sandra geht ans Fenster und winkt zu ihrer Tante hinüber, die bereits im Flugzeug sitzt.	
2	Kurz vor dem Abflug umarmen sich Sandra und ihre Tante noch einmal.		5	Das Flugzeug startet hoch in den Himmel.	
3	Danach checkt Tante Elisabeth ein und läuft zur Startrampe.			*Target (negative)*	

Figure E38. Filler story “Flugzeug” (plane).

no.	Sentence	Picture	no.	Sentence	Picture
1	Maja schlüpft aus ihrem Schlafsack und öffnet leise den Reißverschluss des Zeltes.		4	Auf einmal hört sie hinter sich auf dem Zelt einen Vogel singen.	
2	Während ihre Geschwister noch fest schlafen, setzt sie sich in das Gras hinaus und genießt den ruhigen Morgen.		5	Maja dreht sich um, und der Vogel fliegt hinauf in die Luft.	
3	Völlig entspannt lässt sie die ersten, sanften Sonnenstrahlen in ihr Gesicht scheinen.			*Target (negative)*	

Figure E39. Filler story “Vogel” (bird).

no.	Sentence	Picture	no.	Sentence	Picture
1	Es ist ein besonderer Tag für Familie Schulze, weil Onkel Emil heute seine Hochzeit feiert.		4	Onkel Emil freut sich über den Brief, während der strahlende Karsten der Braut den Blumenstrauß gibt.	
2	Die Mutter hat einen Brief eingesteckt und Karsten darf der Braut einen Blumenstrauß überreichen.		5	Entzückt wirft die Braut den Blumenstrauß in die Höhe.	
3	Als sie ankommen, steht das Brautpaar bereits vor dem Haus und empfängt die Gäste.			*Target (negative)*	

Figure E40. Filler story “Blumenstrauß” (bouquet).

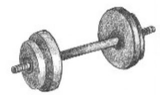
no.	Sentence	Picture	no.	Sentence	Picture
1	Heute kommt Florian in die dritte Klasse, und er ist schon total gespannt.		4	Endlich hat er die Lösung gefunden und ist ganz stolz auf sich.	
2	Herr Becker, der Klassenlehrer, teilt gerade Blätter mit ganz schweren Rechenaufgaben aus.		5	Florian streckt seine Hand hoch.	
3	Als Florian sein Blatt mit den Aufgaben bekommt, stellt er fest, dass er ziemlich lange nachdenken muss, um sie zu lösen.			*Target (negative)*	

Figure E41. Filler story “Hand” (hand).

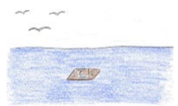
no.	Sentence	Picture	no.	Sentence	Picture
1	Sarah und ihr Bruder Matthias fahren mit einem Floß auf einem großen See.		4	Sie ruft ihrem Bruder zu, damit er sich das Ding genauer ansieht.	
2	Während Matthias rudert, schaut Sarah mit einem Fernglas über die weite Wasserfläche.		5	Da streckt Matthias seinen Arm aus und holt eine Flaschenpost aus dem See herauf.	
3	Dabei sieht sie plötzlich einen rätselhaften Gegenstand, der direkt auf das Boot zu schwimmt.			*Target (negative)*	

Figure E42. Filler story “Flaschenpost” (bottle message).

no.	Sentence	Picture	no.	Sentence	Picture
1	Es ist schönes Wetter, und deshalb gehen Luis und Teresa heute im Park spazieren.		4	Sie streiten so heftig miteinander, dass das Brotstückchen dabei ins Wasser fällt.	
2	Sie haben altes Brot dabei und füttern damit die Enten im Teich.		5	Und während Luis und Teresa sich immer noch streiten, paddelt eine Ente auf dem Teich entlang und schnappt sich das Brotstückchen.	
3	Doch als um das letzte Brotstückchen geht, fangen die Geschwister an sich zu streiten.			*Target (negative)*	

Figure E43. Filler story “Ente” (duck).

no.	Sentence	Picture	no.	Sentence	Picture
1	Gleich geht Katrin zum Schwimmen, und gerade packt sie ihre Badesachen ein.		4	Ob es dort wohl noch etwas Essbares für sie gibt?	
2	Ihre Mutter meint, dass sie unbedingt noch etwas zum Essen mitnehmen muss, denn vom Schwimmen bekommt man ganz schnell Hunger.		5	Doch Katrin hat Glück, denn in der Obstschale liegt noch eine Banane.	
3	Mama hat recht, und so nimmt Katrin ihre Badetasche und geht in die Küche.			*Target (negative)*	

Figure E44. Filler story “Banane” (banana).

no.	Sentence	Picture	no.	Sentence	Picture
1	Heute ist Ostern, und Christoph ist fleißig bei der Ostereiersuche.		4	Es ist ein kleiner Kreisel, und Christoph hebt ihn auf und lässt ihn über dem Tisch los.	
2	Er hat heute schon ganz viele Osternester und Schokohasen entdeckt und ist immer noch nicht müde, weiterzusuchen.		5	Begeistert schaut Christoph zu, wie sich der Kreisel noch lange auf dem Tisch dreht.	
3	Als er nun unter das Sofa schaut, entdeckt er etwas ganz Besonderes.			*Target (negative)*	

Figure E45. Filler story "Kreisel" (spin top).

no.	Sentence	Picture	no.	Sentence	Picture
1	Schon seit Stunden sitzt Armin auf dem Fuß-boden, um ein Puzzle zu lösen.		4	Nun wagt er einen letzten Versuch und tastet sich unter dem Kleiderschrank entlang.	
2	Jetzt fehlt ihm nur noch ein Teil und er ärgert sich, weil er es nirgends finden kann.		5	Und siehe da - mit viel Mühe holt Armin das verlorene Puzzleteil unter dem Schrank hervor!	
3	Außerdem ist es schon spät, und bestimmt kommt Papa gleich herein und sagt Armin, dass er aufräumen soll.			*Target (negative)*	

Figure E46. Filler story "Puzzleteil" (puzzle piece).

no.	Sentence	Picture	no.	Sentence	Picture
1	Ralf macht große Augen, denn er darf heute zum ersten Mal den Werkzeugkasten seines Vaters ausprobieren.		4	Da warnt ihn der Vater, aufzupassen, denn mit dem Hammer kann man sich leicht wehtun!	
2	Der Vater zeigt ihm, wie ein Schraubenschlüssel funktioniert, und Ralf hört aufmerksam zu.		5	Vorsichtig legt Ralf den Hammer in den Werkzeugkasten zurück.	
3	Da greift der neugierige Ralf in den Werkzeugkasten und holt einen Hammer heraus.			*Target (negative)*	

Figure E47. Filler story "Hammer" (hammer).

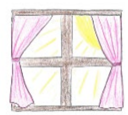
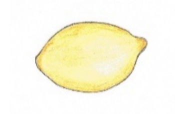
no.	Sentence	Picture	no.	Sentence	Picture
1	Fiona seufzt, denn sie ist immer noch nicht mit ihren Hausaufgaben fertig.		4	Doch in dem Moment fällt ihr ein, dass sie erst übermorgen wieder Deutsch hat.	
2	Die Sonne scheint, und Fiona möchte so gerne draußen im Garten spielen!		5	Erleichtert steckt Fiona das Heft in ihre Schultasche.	
3	Sie schaut in ihr Heft und stellt fest, dass sie noch eine lange Deutsch-Aufgabe vor sich hat.			*Target (negative)*	

Figure E48. Filler story "Heft" (notebook).

no.	Sentence	Picture	no.	Sentence	Picture
1	Die Mutter ruft nach Lars, doch der möchte gerade lieber sein Computerspiel spielen.		4	Als er in den Schrank hineinschaut, stellt Lars fest, dass der Staubsauger viel zu schwer für ihn ist.	
2	Heute ist Großputz angesagt, und Lars soll deshalb staubsaugen.		5	Erleichtert geht Lars zurück in sein Zimmer und spielt weiter.	
3	Widerwillig steht er auf und geht zum Putz-schrank, wo der Staubsauger steht.			*Target (negative)*	

Figure E49. Filler story “Staubsauger” (vacuum cleaner).

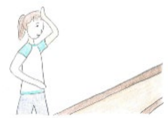
no.	Sentence	Picture	no.	Sentence	Picture
1	Heute regnet es, deshalb gehen Herr und Frau Meyer mit ihrer Tochter Lena zum Bowling.		4	Die Bowlingkugel rollt über die Bahn.	
2	Vor der Kegelbahn stehen viele Menschen, und Lena muss lange warten, bis sie dran kommt.		5	Da gibt es ein lautes Krachen, und Lena freut sich, denn sie hat alle Kegel auf Anhieb umge-worfen.	
3	Als sie nun endlich an der Reihe ist, nimmt sie sich eine Kugel aus dem Korb und wirft.			*Target (negative)*	

Figure E50. Filler story “Bowlingkugel” (bowling ball).

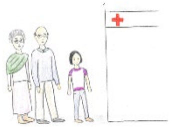
no.	Sentence	Picture	no.	Sentence	Picture
1	Melanie und ihre Großeltern gehen heute ins Krankenhaus, die Mutter besuchen.		4	Das Baby liegt auf der Bettdecke und schaut Melanie an.	
2	Gestern ist nämlich Melanies kleiner Bruder zur Welt gekommen!		5	So klein war ich auch mal, denkt sich Melanie, und schaut lächelnd zurück.	
3	Die Mutter liegt im Bett und freut sich sehr, dass ihre Familie sie besuchen kommt.			*Target (negative)*	

Figure E51. Filler story “Baby” (baby).

no.	Sentence	Picture	no.	Sentence	Picture
1	Paula ist heute alleine in der Wohnung und beginnt, ein Bild zu malen.		4	Als Paula dann endlich den Telefonhörer abnimmt, ist nur noch ein Tuten zu hören.	
2	Doch als sie gerade den ersten Strich macht, hört sie plötzlich das Telefon klingeln.		5	Wer da wohl angerufen hat?	
3	Neugierig öffnet Paula die Tür zum Flur und läuft in die Ecke, aus der das Klingeln kommt.			*Target (negative)*	

Figure E52. Filler story “Telefon” (telephone).

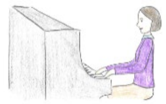
no.	Sentence	Picture	no.	Sentence	Picture
1	Auf dem Weg zur Klavierlehrerin ist Viktoria schon ganz aufgeregt.		4	Lächelnd setzt sich die Lehrerin auf den Hocker und spielt Viktoria ein schönes Lied auf dem Klavier vor.	
2	Heute hat sie nämlich zum ersten Mal Klavierunterricht!		5	Entzückt hört Viktoria zu und wünscht sich, dieses Lied bald selbst spielen zu können.	
3	Als sie dort ankommt, öffnet die Lehrerin ihr die Tür und begrüßt sie freundlich.			*Target (negative)*	

Figure E53. Filler story “Klavierhocker” (piano stool).

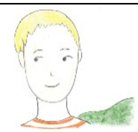
no.	Sentence	Picture	no.	Sentence	Picture
1	Helmut und Jens sind im Garten und spielen Verstecken.		4	Verstohlen blickt sich Helmut um und steckt den Ring auf seinen Finger.	
2	Gerade ist Jens mit Suchen dran, und Helmut hat sich bereits tief im Gebüsch versteckt.		5	Ob Jens ihn jetzt wohl noch finden wird?	
3	Während Jens eifrig nach Helmut sucht, entdeckt der plötzlich einen glänzenden Ring zwischen den Blättern!			*Target (negative)*	

Figure E54. Filler story “Ring” (ring).

Appendix E1.4: Practice trials

no.	Sentence	no.	Sentence
1	Die Mutter ist heute weg, und deshalb hat Julius ganz viel Unordnung im Haus gemacht.	4	Und da kommt auch schon Mamas Auto angefahren und stellt sich auf den Parkplatz vor dem Haus.
2	Erst am späten Nachmittag fällt ihm ein, dass er noch aufräumen muss, bevor Mama zurückkommt.	5	Kurze Zeit später öffnet Mama die Haustür.
3	Ängstlich läuft Julius zum Fenster und schaut hinaus.		*Target (positive)* 

Figure E55. Practice story “Auto” (car).

no.	Sentence	no.	Sentence
1	Alex und seine Schwester Emily spielen Fußball auf dem Hof.	4	Doch da bemerkt Alex, dass er den Ball viel zu weit geschossen hat!
2	Alex steht im Tor und freut sich, weil er den Schuss von Emily gehalten hat.	5	Emily versucht, hinterherzurennen, doch da rollt der Fußball bereits auf die Straße.
3	Jetzt nimmt er Anlauf und macht einen kräftigen Abstoß.		*Target (positive)* 

Figure E56. Practice story “Fußball” (football).

no.	Sentence
1	Es hat gerade geklingelt, und Meike und die anderen Kinder rennen ins Klassenzimmer.
2	Als Meike sich auf ihren Platz setzt, traut sie ihren Augen kaum.
3	Auf ihrem Tisch liegt eine schöne Blume!
4	Wer sie dort wohl hingelegt hat?
5	Meike nimmt die Blume in die Hand und riecht daran.
	Target (positive) 

Figure E57. Practice story "Blume" (flower).

no.	Sentence
1	Tim freut sich auf die Schule, denn heute ist Sport.
2	Deshalb packt er auch schon eifrig sein T-Shirt und seine kurze Hose ein.
3	Jetzt fehlen nur noch die Turnschuhe.
4	Tim überlegt kurz und öffnet dann die Wohnungstür, um hinauszusehen.
5	Und tatsächlich - die Turnschuhe stehen vor der Tür.
	Target (negative) 

Figure E58. Practice story "Turnschuhe" (sneakers).

no.	Sentence	no.	Sentence
1	Heute wird im Kindergarten gemalt, doch Nora freut sich überhaupt nicht darauf.	4	In dem Moment bückt sich Kim und hebt etwas vom Boden auf.
2	Sie muss nämlich ausgerechnet neben Kim sitzen, und die beiden mögen sich gar nicht.	5	Lächelnd dreht sie sich zu Nora hinüber und gibt ihr den Stift in die Hand.
3	Als Nora nun missmutig mit dem Malen beginnen will, stellt sie plötzlich fest, dass ihr Stift nicht mehr da ist.		*Target (negative)* 

Figure E59. Practice story "Malstift" (color pencil).

no.	Sentence	no.	Sentence
1	Lisa ist stolz, denn heute darf sie zum ersten Mal alleine mit dem Hund Gassi gehen.	4	Sanft streichelt Lisa den Hund und legt ihm vorsichtig die Leine an.
2	Sie geht zur Garderobe und holt die Leine aus der Schublade.	5	Lisa öffnet die Haustür - und schon geht's los!
3	Dann ruft sie und wartet, bis der Hund zu ihr kommt.		*Target (negative)* 

Figure E60. Practice story "Hund" (dog).

Appendix E2: Instruction

Original instruction	English translation
Schön, dass du heute hier bist und mitmachst! Du bekommst gleich Geschichten vorgelesen, bei denen du ganz genau aufpassen musst. Denn immer wenn die Geschichte vorbei ist, kommt ein kurzer Piepston, und dann bekommst du ein Bild zu sehen, zum Beispiel eine Blume oder ein Malstift.	Nice to have you here taking part! You'll get two tasks. You're going to listen to stories and you have to pay much attention to them. Because whenever the story ends, there's a short beep, and then you see a picture, a flower for example or a color pen.
Dann musst du ganz schnell rausfinden, ob das, was du da siehst, in der Geschichte vorgekommen ist. Wenn ja, dann drückst du auf den glücklichen Smiley. Wenn nein, dann drückst du auf den traurigen Smiley. Du darfst nicht zu lange warten, sonst ist das Bild weg und deine Antwort zählt nicht mehr.	Then you have to get very fast if that what you see happened in the story. If so, then you press the happy smiley. If not, you press the sad smiley. You should not wait too long because otherwise, the picture is gone and your answer doesn't count anymore.
Am besten, du legst immer den rechten Zeigefinger auf den traurigen Smiley und den rechten Mittelfinger auf den glücklichen Smiley *[Versuchsleiter*in macht Antwortverhalten vor und lässt das Kind nachmachen].*	So your best bet is to put the right index just over the sad smiley and the right middle finger over the happy smiley *[Experimenter shows response behavior and waits until the child imitates it].*
Genau so! Wenn du bereit bist, dann machen wir erst mal ein paar Übungsgeschichten zusammen. Dazu setzt du bitte den Kopfhörer auf. Achtung, los geht's! *(Versuchsleiter*in drückt Enter)*	Exactly! When you're ready now, let's start with some practice stories. So please take now your headphones on. Attention… here you go! *(Experimenter presses Enter)*
Es folgen die 6 Übungsaufgaben.	*The 6 practice trials follow.*
Alles klar soweit? Bitte denk daran: Immer die Finger an den Tasten lassen und wenn du das Bild siehst und weißt, ob es in der Geschichte vorkam, ganz schnell drücken! Hast du noch Fragen? *(Versuchsleiter*in beantwortet ggf. Fragen und drückt danach Enter)*	Are you ready? Please remember to have always the fingers over the keys, and whenever you see the picture and know if it has happened in the story, press quickly! Have you got any questions? *(Experimenter answers questions, if any, then presses enter)*
*Experiment beginnt. Versuchsleiter*in erinnert das Kind bei Bedarf an die vorgesehene Fingerstellung.*	*Experiment starts. Experimenter reminds the child of the designated positions of the fingers.*