Title:
Online emotional inferences in written and auditory texts – a study with children and adults

Authors:
Anna Katharina Diergarten (University of Würzburg)
Gerhild Nieding (University of Würzburg)

Abstract

Emotional inferences are conclusions that a reader draws about the emotional state of a story's protagonist. In this study, we examined whether children and adults draw emotional inferences while reading short stories or listening to an aural presentation of short stories. We used an online method that assesses inferences during reading with a reaction time paradigm. Children aged 8 and 10 years and adults took part. We varied whether the short stories emphasized a certain goal of the protagonist in order to proof our assumption that protagonist-goals in stories help readers to build emotional inferences. Additionally, we assessed the updating capacity of our participants' working memories assuming a positive influence. Results reveal that participants of all age groups drew emotional inferences. Type of text presentation, goal-emphasis, and updating capacity influence whether emotional inferences are built and how precise these inferences are. The way in which these variables influenced emotional inferences was moderated by the age of the participants.

KEYWORDS: Children, development, emotional inferences, reading comprehension, listening comprehension, working memory

Introduction

It is a widely recognized fact that reading requires more than just decoding letters and putting meaning to words. Although these basic language skills are no doubt important for early reading development, another important aspect is the development of narrative comprehension skills. In order to comprehend a narrative, readers form a mental representation of the text on three levels (van Dijk & Kintsch, 1983). Two of these levels are based directly on the text: the surface form—that is, the exact words and syntax of the text—and the text base, which represents the propositional structure of the text. The third and highest level goes beyond the mere representation of the text itself, as it also includes a representation of the situation described by the text, a so-called situation model (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). In this situation model, which is essential for narrative comprehension (Nieding & Ohler, 2008; Oakhill & Cain, 2004), readers represent the people, setting, actions, and events of the text, which are either explicitly mentioned in the text or filled in by the reader through various kinds of inferences (Orrantia, Múñez, & Tarín, 2014; Zwaan, Langston, & Graesser, 1995).

The mental representations in the situation model are organized along five dimensions: space (spatial aspects of the situation), time (information about the situation's time frame), protagonist (what persons are involved), causality (causal connection with prior events), and intentionality (goals; e.g., Therriault & Rinck, 2007; Zwaan & Radvansky, 1998). These dimensions are monitored during reading, and shifts in any of the dimensions are processed and incorporated into the representation (Zwaan, Langston, & Graesser, 1995; Zwaan, Magliano, & Graesser, 1995). Although there is empirical evidence for all five dimensions (Therriault & Rinck, 2007; Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998), there are different opinions concerning their order of dominance. Some researchers assume that all dimensions are equally weighted in situation model construction (e.g., Zwaan, Langston, & Graesser, 1995), while others claim the superiority of one or several dimensions (e.g., Rinck & Weber, 2003). Many researchers agree that intentionality and causality are most important for

story comprehension (Zwaan & Radvansky, 1998). Causal inferences connect causes and consequences across events, whereas intentionality inferences provide the motivation for actions throughout the narrative (Tompkins, Guo, & Justice, 2013) and thus inferences in both of these dimensions help readers in their "effort after meaning" (Graesser, Bertus, & Magliano, 1995). This is especially important for children's narrative comprehension, because children often do not connect information that is separated within the text (Ackerman, 1984; Ackerman & McGraw, 1991; Orrantia et al., 2014). That is, they are primarily concerned with establishing local coherence, making sense between events closely following one another. Goals within narrative structures help readers to focus on the global structure of narratives (Unsöld, 2008). In addition, several studies have demonstrated that goal inferences are more important than other kinds of inferences in predicting narrative comprehension (Kendeou, Bohn-Gettler, White, & van den Broek, 2008; Lynch & van den Broek, 2007). For instance, Tompkins et al. (2013) counted the number of different kinds of inferences that 4- to 5-year-old children spontaneously made while narrating a wordless picture book. The number of goal inferences was significantly related to their story comprehension score.

In the current study, we focused on emotional inferences—that is, inferences about the protagonist's emotional state.

Emotional Inferences

The typical way of assessing emotional and other inferences is the reading-time paradigm: Gernsbacher, Goldsmith, and Robertson (1992) confronted their participants with stories containing specific emotional themes without explicitly mentioning the intended emotion until the final sentence. They presented the story sentence by sentence on a computer screen. The final sentence was the target sentence and included the mentioning of the protagonist's emotion, for example, "it would be weeks before Tom's guilt would subside" (Gernsbacher et al., 1992, p. 93). Participants were asked to press a key as soon as they had finished reading the sentence. Reading times were faster when the emotion stated in the final

sentence matched the emotion that was implied by the story than when the emotion did not match. As the protagonist's emotion was not mentioned before, the so-called mismatch effect is interpreted as a confirmation that the reader has inferred the protagonist's emotion from the story.

Further research focused on the specificity of emotional inferences. For instance, the sentence "Peter is smiling" only allows the reader to infer that Peter is experiencing a pleasant feeling, but not whether he is, for instance, happy or proud. Further information about Peter's current situation would be necessary to make a specific emotional inference. In fact, the specificity of emotional inferences depends on the amount of background information available (Molinari et al., 2009) as well as on the reader's emotional knowledge (Gernsbacher & Robertson, 1992; Diergarten & Nieding, 2015). Readers do not differentiate between very similar emotional terms such as "depressed" and "miserable" (Gygax, Oakhill, & Garnham, 2003), but between less similar emotions with the same valence, such as pride and happiness (Ohler, Nieding, & Töpper, 2002; Diergarten & Nieding, 2015).

The method of reading latencies described above is an example of an online measure, an instrument that assesses inferences built during an uninterrupted comprehension process. Offline measures, on the other hand, assess inferences after the comprehension process is finished, for example, by directly asking about the protagonist's emotion. The problem with data collected by offline measures is that they do not mirror the processes that occur during the spontaneous processing of narratives. For that reason, one cannot exclude the possibility that participants only thought about the protagonist's emotion because they were asked about it later, but wouldn't have thought about it under more natural reading circumstances. Thus, data collected with offline measures only indicate what participants can do when they are asked to, but not necessarily what they would do spontaneously while reading on their own (Kendeou et al., 2008).

Despite this problem, until recently there were hardly any studies using online measures to test emotional inferences in children. The main reason for this is that traditional online methods, such as the reading-time paradigm, require fluent reading skills and are therefore not applicable to young children. Several studies using offline measures have found that even preschool children can infer a protagonist's emotional state when they are asked to do so. Also, Tompkins et al. (2013) found that children built several kinds of inferences spontaneously while narrating a wordless book, including emotional inferences.

To examine the nature of children's online emotional inferences in more detail, we developed a reaction-time methodology (Diergarten & Nieding, 2015). We conducted two studies with 5-, 8-, and 10-year-old children and adults. The first study used televised narratives, the second audio stories. The results indicated that even 5-year-old children infer the protagonist's emotions spontaneously, both in aural and televised narratives. A developmental trend occurred concerning the specificity of emotional inferences. Younger children only inferred the emotional valence of the protagonist's feelings (e.g., feels good vs. does not feel good) whereas 10-year-old children and adults inferred the exact emotional state (e.g., happy vs. proud).

In the current study, we examined whether children also build emotional inferences during reading or when someone is reading to them. Although our previous study also included an aural condition, this differed from the one that we designed for the current study. In the previous study, the aural condition was designed to be comparable to audio plays, and therefore it contained music and different voices for the different characters. In the current study, the aural condition only contained a recorded reading of a short story.

Comprehension of Written and Aural Texts

Years before children learn how to read, they become acquainted with listening to and trying to understand stories that others read or tell to them. There are two different views regarding the question of whether these early acquired aural comprehension skills (also referred

to as *auding*; see Brown, 1954; Stitch & James, 1984) are the same processes that are later relevant for reading comprehension. The *unitary process* view assumes that the same cognitive processes underlie auding and reading comprehension. Scientific support for this view comes from correlational and longitudinal studies that discovered relations between these two processes. For instance, Kendeou et al. (2005; 2008) found that narrative comprehension of aural narratives at the age of 6 directly predicted children's reading comprehension two years later. The *simple view of reading* model is also in line with the unitary process view. It postulates that reading comprehension is determined by two components: the decoding of words and listening comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990). The model predicts that in the first years of reading instruction, decoding capability is the strongest predictor of reading comprehension, and that the influence of listening comprehension on reading comprehension becomes stronger as decoding capability develops. This developmental trend, however, seems to be true only for the learning process of languages with complex orthographies (such as English), while in transparent orthographies (such as German) listening comprehension is an important predictor for reading comprehension from the very early stages of reading development (Florit & Cain, 2011).

The *dual process view*, on the other hand, claims that different processing strategies and memory processes occur during reading or auding and that different mental representations are built depending on text modality (Kürschner, Schnotz, & Eid, 2006). Support for this view comes mainly from studies with adult participants, which typically find that the comprehension of written texts is superior to the comprehension of auditory texts (Kürschner et al., 2006; Zumbach & Schwartz, 2014). However, some results also show that the advantageous presentation form depends on a person's preferences for reading vs. listening (Kürschner, Schnotz, Eid, & Hauck, 2005) and on the type of task. Research has revealed that readers pay more attention to detail and therefore produce better results concerning detail recall (Rubin, Hafer, & Arata, 2000; Vidal, 2011). Listeners, on the other hand, try to build a global

representation, that is, a situation model of the narrative, which leads to better results concerning comprehension (Hildyard & Olson, 1978). This effect might be due to listeners' inability to go back in the text to look up certain details, thus applying the strategy of focusing on the gist and the overarching meaning of the text. However, Hildyard and Olson conducted their study with children (10- and 12-year-olds), so this effect might also be due to the developmentally typical superiority of auding. Stitch and James (1984) published a model that, based on 31 studies comparing auding and reading, depicts the grade level at which auding or reading performance is better or equal. According to this model, up to the seventh grade, auding is superior to reading, followed by a long period during which auding and reading performance are equal, until reading ability surpasses auding ability around the 12th grade. An explanation for these findings is that it takes years of practice for reading to become automatic. Once this has been established, students have a lot more experience with learning from written texts, as learning difficult material through auding is hardly ever required in or outside of school (Stitch & James, 1984).

Text Comprehension and Working Memory

Whether reading or auding, text comprehension requires information to be kept in mind that is relevant to interpreting information given later in the text (Palladino, Cornoldi, Beni, & Pazzaglia, 2001). Low working memory capacity has been identified as a reason for poor text comprehension in several studies (e.g., Cain, Oakhill, & Lemmon, 2004; Oakhill, Hartt, & Samols, 2005). This finding is easily explained by the fact that the reader needs to keep in mind information that is relevant for interpreting information given later in the text (Palladino et al., 2001). However, working memory capacity as it is traditionally operationalized, such as in span tests that measure the number of items that can be held in the mind over a period of time, is related to tests that measure text comprehension at the text base level, but not on situation model tests (Radvansky & Copeland, 2004; Radvansky, Gibson, & McNerney, 2014). Radvansky and Copeland (2004) explain this finding with the fact that processing at the situation model level

goes beyond simply keeping information available. Instead, in order to maintain a coherent situation model, memory content needs to be modified constantly to assimilate new information, that is, substituting new information for information no longer relevant (Morris & Jones, 1990). This process is called updating (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Radvansky and Copeland (2001) found that updating ability was related to situation model processing. Several studies have demonstrated that updating capacity is a relevant factor when comparing good and poor comprehenders, in both younger students (Carretti, Cornoldi, Beni, & Romano, 2005) and older students and young adults (Palladino et al., 2001). Carretti et al. (2005) found that a particular type of error in their updating task, called *delayed intrusion error*, turned out to be the most relevant criterion for discriminating between good and poor comprehenders. This error represents the failure to inhibit information that was relevant initially but which later becomes irrelevant.

Purpose and Design of the Present Research

Our own previous research (Diergarten & Nieding, 2015) revealed that children (5, 8, and 10 years old) and adults build emotional inferences online when listening to audiobooks or watching television. The goal of the present research was to examine whether they also infer the protagonist's emotional state when they are reading or listening to short narratives. We chose the same age groups as in our prior research except for 5-year-olds, as most of them are not yet able to read. We adapted our methodology to test for online emotional inferences in written narratives. For that purpose, we created short stories in which different child protagonists experienced one of four emotions: joy, pride, sadness, or anger. After each story, a male voice added the last sentence to the story by stating how the protagonist felt. This statement was either consistent with the described situation or inconsistent. Participants reacted to the last sentence by pressing one of two keys. As in the reading-time paradigm, reaction times to consistent and inconsistent last sentences were compared, and shorter times for consistent sentences interpreted as an indicator of emotional inferences.

To test whether the ability to build emotional inferences depends on the type of text presentation, half of the stories were presented aurally, that is, as pre-recorded readings of the texts, and the other half as written texts. As outlined in the introduction, there are different viewpoints as to whether reading and auding require the same or different mental processes. An important factor in this discussion is age. A merit of our study, therefore, is that we tested both children and adults and can thus gain information about the differences between reading and auding for emotional inferences in different developmental stages.

The simple view of reading model predicts that in the first years of reading instruction, reading comprehension is mainly influenced by children's word decoding capability, and that the influence of listening comprehension on reading comprehension becomes stronger once decoding becomes more automatic. The gap between reading and auding ability closes around seventh grade (Stitch & James, 1984) and eventually, with the beginning of adulthood, reading comprehension becomes superior to auding (Kürschner et al., 2006; Rickheit & Strohner, 1983; Rubin et al., 2000). Based on this research, we expect that the 8- and 10-year-olds in our study should be better at building emotional inferences in the auding condition, while adults should profit from the reading condition.

Not only the type of text presentation but also the structures of narratives influence the focus of readers. Goals in narratives, especially, are known to lead readers' focus on the global theme of a text (e.g., Unsöld, 2008). We assumed that this should be true for emotional inferences as well because the outcome of goals is an important source for emotions. In order to test this assumption, we created two versions of every story, one of which explicitly mentioned and explained the protagonist's goal and thereby his or her motivation. We assumed that the story versions with a focus on a specific goal would facilitate the building of emotional inferences more readily compared to the story versions without a goal.

Additionally, we assessed the updating capacity of participants' working memory to gain further information as to which individual skills influence emotional inferences. As far as

we know, neither working memory capacity in general nor updating capacity specifically have been examined in relation to emotional inferences. However, updating capacity seems to be especially important in building situation models, that is, when trying to connect new information with old information and trying to establish global coherence. For this reason, we assume that updating capacity is important for emotional inference and expect that participants with relatively high updating capacity will infer the protagonists' emotions better than participants with comparatively low updating capacity. We expect this effect especially in the two groups of children, due to their age-related normal limitations in working memory capacity (Cowan et al., 2003). In addition, Carretti et al. (2005) argue that working memory is more important for children's narrative comprehension because adults are better able to rely on their general knowledge.

Method

Participants

Ninety-eight participants took part in our study. The data on two of the participants had to be excluded later due to missing data (see the section "Processing of Data" below). This left 96 participants, with 32 in each age group.

Children were second and fourth graders recruited from a primary school located in a suburb of Würzburg, Germany. The school is attended by children from middle- and upper-middle-class families. There were 16 boys and 16 girls in each group. Their mean age was 7;11 (years; months, *SD* = 0;4) and 10;0 (*SD* = 0;5). The adults were university students majoring in psychology, 19 of whom were female and 13 male. Their mean age was 22;3 years (*SD* = 2;11). German was the native language of all participants.

Material

Emotional inferences. The test was based on our measurement for emotional inferences in films (Diergarten & Nieding, 2015). Whereas in this prior study we presented film clips and auditory texts, in the present study we presented written and auditory texts. We

created 24 short stories about children's everyday life experiences. In each of the stories, a child protagonist feels a certain emotion that is not explicitly mentioned in the text, but which is established by the circumstances of the story. Six stories each were about joy (for example, the protagonist visits the zoo and is lucky because the lions are not sleeping behind the big stones but walking through their compound, so he can take a picture), pride (the protagonist wins a long-jump contest), sadness (the protagonist's hamster dies), and anger (a boy destroys the protagonist's sand castle deliberately).[1] An example story is printed in Table 1. For each emotion, half of the stories had a male and the other half a female protagonist.

In writing the emotional stories, we were guided by the emotional appraisal model (Roseman, Antoniou, & Jose, 1996). This model displays how different emotions arise depending on the person's appraisal of events concerning one's motives, that is, whether they are appetitive or aversive and whether the event is consistent or inconsistent with one's motives. We made sure that the stories were easy to read by using short sentences and simple words. We aimed to make the texts similar to second-grade reading material.

Pilot study. We conducted a pilot study to certify that our short stories were unambiguous concerning the intended emotion. A total of 28 university students majoring in psychology (18 female, 10 male) with a mean age of 22;0 years (*SD* = 3;7 years) completed the pilot study. None of these participants took part in our main experiment. We asked participants to state the emotion the protagonist of each story felt, allowing them to state several emotions if they thought the emotional state was ambiguous. We counted how many times each emotion was mentioned in every story. Stories for which at least one participant had stated two emotions (for example "joy and pride") were excluded. Out of the remaining stories, we chose the four stories for each emotion for which most participants had stated the emotion that was intended by us or a previously defined synonym (for example, "happiness" instead of "joy"). The correct

[1] The German emotion words were "Freude," "Stolz," "Trauer," and "Wut."

rating in the 16 stories used in the main study was between 71% and 96%. The highest scores were for the four stories about pride (ranging from 86% to 96%, *M* = 91.85, *SD* = 5.23), while the lowest were for anger (ranging from 71% to 86%, *M* = 78.55, *SD* = 8.25). The results for joy and sadness were very similar, both ranging from 75% to 96% (joy: *M* = 86.6, *SD* = 8.92; sadness: *M* = 87.5, *SD* = 9.45).

Written and auditory presentation. The average length of the stories that were chosen after the pilot study was 73.78 words (*SD* = 8.60) and 7.72 sentences (*SD* = 1.51) per story. The average word length was 4.94 letters (*SD* = 0.21) or 1.59 syllables (*SD* = 0.06). The stories were presented either in print or in an audio version. We varied the presentation in a within-design; that is, each participant read eight stories in the reading condition and heard eight stories in the listening condition.

In the reading condition, the texts were presented on a computer screen. We used the character font "Century Gothic." The font size was 32. The texts were aligned left and printed in white letters on a black background. The whole text was presented at the same time in a single block. We asked the participants to read the story out loud in order to make sure they did not skip any words or sentences. We captured the reading times by having the experimenter press a button as soon as the participant finished reading. We only assessed reading times to screen for very slow readers, so a more accurate measurement was not necessary. After the experimenter pressed the button, the text disappeared from the screen and the reaction-time task started (see below).

For the audio version, we presented a previously recorded reading of the stories by a male adult. He modulated his reading concerning emphasis, volume, speed, and tone. The mean duration of the auditory texts was 29.94 seconds (*SD* = 3.60). The screen remained black during the auditory presentation.

Goal emphasis. We had two versions of every story and varied these in a within-design; that is, every participant was presented with eight stories with and eight stories without goal

emphasis. The beginning and end of the stories were identical in both versions; just two sentences in the middle of the stories were varied between versions. In the goal-emphasis version, the two middle sentences specifically elaborated on the protagonist's goal, while the two middle sentences in the version without goal emphasis were filler sentences that made sense in the story but did not relate to the goal. For instance, in the story in which the child (Thomas) sees a lion in the zoo, the middle sentences in the goal-emphasis version describe that Thomas is supposed to write a report about lions for school and therefore wants to take a picture of them. The filler sentences in the version without goal emphasis describe how Thomas and his father visit other animal compounds in the zoo before they go to see the lions. Another example of a story in both versions is given in Table 1.

Reaction-time task. After each story, a blue dot appeared in the middle of the screen and a male voice (the same voice that read the stories in the listening condition) spoke the target sentence. Note that we always presented the target sentence aurally—in both conditions—in order to avoid the influence of individual differences in reading speed. The target sentence was "[name of the protagonist] feels [one of the four emotions: sad, angry, proud, or happy]." For example, "Thomas feels happy." This emotion either matched or did not match the emotion that the protagonist of the story actually felt, and the mismatching emotions were either of the same or of the opposite emotional valence of the actual emotion, resulting in three conditions of congruency. For example, in the story where Thomas gets to take a picture of the lions, the spoken sentence "Thomas feels happy" is the congruent condition (C), "Thomas feels proud" is the incongruent with same valence condition (I-SV), and "Thomas feels sad" and "Thomas feels angry" are the incongruent with opposite valence conditions (I-OV).

After the target sentence was presented, the participants' reaction-time task was to press one of two buttons as quickly as possible. If the male voice had stated a positive emotion, they were supposed to press a "smiley" key (☺), if it stated a negative emotion the "frowny" key (☹).

After the participant pressed one of the two keys, the experimenter asked a question about the story which assessed general comprehension of the story. The purpose of the questions was to motivate participants to pay attention to the story's content. They were not important for data analysis.

– Enter Table 1 about here –

Updating capacity. The updating task was based on Carretti et al. (2005). We varied their procedure slightly by extending the number of pictures and words from 15 to 25 in order to account for the adults that took part in our study and in order to avoid ceiling effects. Participants saw a column of 25 pictures (line drawings) and the experimenter read a list of 25 words. Some of the words were represented as pictures in the column; others were not. The participants' task was to recall the five highest (or lowest) pictures in the column that were named by the experimenter. In order to solve this task, participants needed to carefully screen the column of pictures, recall which pictures were named by the experimenter and—most importantly—constantly update the information about which of the named pictures were the highest or lowest five (see Carretti et al., 2005, p. 53, for a depiction of the procedure).

Ten words in the list were filler words, that is, abstract nouns such as "feeling" or "luck" that were clearly not associated with any picture, so participants didn't need to consider these at any time. Five words of the list were the correct words (*correct recall*), five were words that could be excluded immediately, because they did not fulfill the criterion of the task (*immediate intrusion errors*), and five words had to be considered and remembered for a while and then be excluded later because they no longer fulfilled the criterion of being among the highest or lowest five pictures. The latter five words are called "*delayed intrusion errors"* and were essential for the total score.

Procedure of the updating task. Before the actual task, the experimenter showed the participant two sheets on which were represented all of the 200 pictures that were part of the updating task. The participant was asked to name every picture in order to make sure that all

pictures were familiar to them. If necessary, the experimenter named a picture, or explained that a different word for the picture would be used in the experiment. For instance, if the participant said "puppy," the experimenter would explain that they would use the word "dog" in the task. After this procedure, the experimenter explained the task by presenting the first column as a practice trial. The experimenter emphasized that the participant was supposed to use his or her eyes only and not, for example, point to the pictures. Once the participant understood the task, the experimenter read the list of words for the practice trial. If the participant solved the practice trial correctly (i.e., named the five highest-named pictures) the experimenter started the first experimental trial. Otherwise, he explained the task again until it was fully understood. In the first practice trial and the first three experimental trials, the five highest pictures were the target. Afterwards, another practice trial was conducted to switch the target to the five lowest pictures. After the participant understood the new challenge, three experimental trials with the lowest five target pictures followed. The experimenter wrote down the participants' answers.

Processing the updating task data. The number of delayed intrusion errors was counted. For each experimental trial, a maximum of five delayed intrusion errors was possible, so the total score ranged from 0 to 30 points. We also counted the other types of errors (omissions, immediate intrusion errors, and others) and the correct answers, but will not discuss them in detail.

Design and Procedure

All participants were tested in individual sessions in a quiet room at their school or university. Children were tested in two sessions on different days. The updating task was conducted in the first session and took about 15 minutes. The reaction-time task in the second session took about 20 to 40 minutes, depending on the child's reading times. The lapse of time between the two sessions ranged from six to 24 days (M = 13.17, SD = 3.75). Adults were tested in one single session with an average duration of 35 minutes.

Emotional Inferences

Each participant was presented with all 16 stories. However, there were eight conditions that counterbalanced the presentation mode and the presented version of the stories, so that each participant was presented with eight written and eight auditory texts, four of each being goal-emphatic or non-goal-emphatic. This resulted in a 2 (presentation mode: written vs. auditory) x 2 (goal: with vs. without goal emphasis) x 3 (congruency: congruent [C], incongruent—same valence [I-SV], incongruent—opposite valence [I-OV]) x 3 (age) design, with presentation mode, goal emphasis, and congruency as within-subject variables. We also varied the sequence in which the stories were presented. There were four different sequences, each of which was used for two conditions. In those sequences, we made sure that no more than two stories with the same emotion were presented consecutively. Written and auditory texts were always presented alternately.

Additionally, we varied the arrangement of the smiley and frowny keys, that is, which was on the right or left side of the keyboard. Both the story sequence and the arrangement of the keys were provisions to prevent sequence or handedness effects, but were not included as variables in the data analysis.

Four participants of each age group were assigned to each of the eight conditions. In the two groups of children, we had an equal number of boys and girls, so we assigned two boys and two girls to each condition. In the case of the adults, we had more female than male participants, so three conditions had three female and one male participant. Aside from gender, the assignment to the conditions was random after order of participation.

Processing of data. The program Presentation® collected information about the participants' reading times, their reaction times, and whether they pressed the correct or false key. In order to screen for extreme reaction times, we defined a criterion of three times the standard deviation higher or lower than the mean reaction time of the respective age group (see Rey, 2012). All extreme values were deleted. Response times for incorrect responses were

deleted as well. Subjects with a high percentage of missing data (more than 25%, or more than four responses) were excluded from further analysis. The data of two 10-year-old children were excluded for this reason. Missing values of the remaining participants were filled in by the mean value of the respective age group and condition.

In order to screen for slow or fast readers, we used the same criterion of three standard deviations higher or lower than the mean of each age group's reading times. None of the participants' reading times were extreme according to this criterion.

Results

Descriptive Statistics

– Enter Table 2 about here –

Table 2 displays the mean scores of the three age groups for errors in the reaction-time tasks, reaction times, reading times in the reading condition, and the story comprehension questions. The very low number of errors confirms our previous findings (Diergarten & Nieding, 2015) that this reaction-time task is fairly easy to do for children and adults. The high means of the comprehension questions reveal that the stories were easy to comprehend.

A MANOVA was calculated with the above-mentioned four scores as dependent variables and age as a between factor. The age group did not have a significant influence on the number of errors ($F < 1$). However, it did have a significant influence on all other variables: reaction times: $F(2, 93) = 66.42$, $p < .001$, $\eta_p^2 = .59$; reading times: $F(2, 93) = 89.97$, $p < .001$, $\eta_p^2 = .66$; comprehension questions: $F(2, 93) = 14.32$, $p < .001$, $\eta_p^2 = .24$.

Scheffé post hoc tests showed that for reaction and reading times, the differences between all age groups was significant ($p < .001$ in all tests). In the case of the comprehension questions, the result was due to the difference between the 8-year-olds and both older age groups (10-year-olds: $p = .001$; adults: $p < .001$), whereas the difference between 10-year-olds and adults was not significant ($p = .591$).

Online Emotional Inferences

To test whether participants built online emotional inferences and whether this was influenced by the variations in medium and goal emphasis, we calculated a multifactorial ANOVA. Because our hypotheses were different for children and adults, we conducted the analyses separately for each age group. Thus, we calculated three four-factorial ANOVAs with the three types of reaction times (C, I-SV, I-OV), the two types of media (reading vs. listening), and the two goal conditions (goal emphasis vs. no goal emphasis) as within-subject factors.

The results of the 8-year-olds showed a significant main effect of the type of reaction time, $F(2, 62) = 6.46$, $p = .003$, $\eta_p^2 = .17$; however, neither the main effect of the type of medium or the main effect of the goal condition nor any interaction effect was significant (all $F < 1$). Concerning the main effect of the type of reaction time, contrasts revealed that reaction times were significantly shorter in the congruent than in the I-OV condition, indicating inferences of the emotional valence, $F(1, 31) = 12.78$, $p = .001$, $\eta_p^2 = .29$, while the difference between C and I-SV was not significant, $F(1, 31) = 1.11$, $p = .300$, $\eta_p^2 = .04$, which means that the 8-year-olds did not build exact emotional inferences. Figure 1 demonstrates the differences in the three types of reaction times.

– Enter Figure 1 about here –

In the 10-year-olds' data, the main effect of type of reaction time was not significant, $F(2, 62) = 2.84$, $p = .066$, $\eta_p^2 = .08$. Figure 1 shows that the differences between C, I-SV, and I-OV are very small. The other two main effects were not significant either: type of reaction time, $F(2, 62) = 2.84$, $p = .066$, $\eta_p^2 = .08$, and type of medium and goal condition, $F < 1$ each. However, there was a significant interaction between reaction times and medium: $F(2, 62) = 4.57$, $p = .014$, $\eta_p^2 = .13$. The other interactions were not significant: goal emphasis x reaction times, $F(2, 62) = 2.49$, $p = .091$, $\eta_p^2 = .07$; both other interactions, $F < 1$. To break down the interaction effect between medium and reaction times, we calculated separate *t*-tests for the reading and the listening conditions (see Table 3 for the reaction times in the different

conditions and the results of the *t*-tests). Results showed that the only significant reaction-time difference was in the reading condition between C and I-SV, indicating exact inferences, whereas 10-year-olds did not build any emotional inferences in the listening condition.

– Enter Table 3 about here –

In the adults' group, the main effect of the reaction times was not significant ($F < 1$), because the three types of reaction times hardly differed from one another (see Figure 1). The main effects of medium and goal emphasis were both significant, which goes back to shorter reaction times in the reading condition and in the goal-emphasis condition, respectively: medium, $F(1, 31) = 4.98$, $p = .033$, $\eta_p^2 = .14$; $M_{\text{reading}} = 706$ ($SD = 114$), $M_{\text{listening}} = 735$ ($SD = 119$); goal emphasis, $F(1, 31) = 5.25$, $p = .029$, $\eta_p^2 = .15$; $M_{\text{goal emphasis}} = 713$ ($SD = 114$), $M_{\text{no goal emphasis}} = 728$ ($SD = 115$).

There were significant interactions between reaction times and medium, $F(2, 62) = 3.94$, $p = .025$, $\eta_p^2 = .11$, and between reaction times and goal-emphasis condition, $F(2, 62) = 8.32$, $p = .001$, $\eta_p^2 = .21$. The three-way interaction was not significant, $F(2, 62) = 3.08$, $p = .053$, $\eta_p^2 = .09$. In order to break down the interaction effects (reaction times x medium and reaction times x goal emphasis) we calculated several *t*-tests for these conditions (see Tables 3 and 4 for the reaction times in the different conditions and the results of the *t*-tests). Concerning type of medium, these *t*-tests revealed no significant reaction-time difference in the listening condition, but a significant difference in the reading condition between C and I-SV, indicating exact emotional inferences in the reading condition. Concerning goals, in the goal-emphasis condition there was a significant difference between C and I-SV, indicating exact emotional inferences. The difference between C and I-OV was not significant. In the no-goal-emphasis condition, although the differences between reaction times were significant, these differences do not indicate emotional inferences because the difference was in the other direction, the incongruent reaction times being shorter than the congruent ones.

– Enter Table 4 about here –

In summary, the emotional inferences of the 8-year-olds were uninfluenced by our two variations. The inferences of 10-year-olds and adults were better under the reading condition. Additionally, adults profited from the goal-emphasis condition.

In the following section, we will report the results concerning the updating task and the possible influences on emotional inferences.

Influence of Updating Capacity on Inference Generation

The relatively high scores of correctly recalled items in all age groups (see Table 5) indicate that the higher number of pictures and words that we used compared to Carretti et al. (2005) was appropriate.

– Enter Table 5 about here –

Both the age effect for correctly recalled items, $F(2, 93) = 11.14$, $p < .001$, $\eta_p^2 = .19$, and that for delayed intrusion errors, $F(2, 93) = 11.64$, $p < .001$, $\eta_p^2 = .20$, were significant. Scheffé post hoc tests revealed that both effects are due to significant differences between 8-year-olds and the two older age groups, whereas the differences between 10-year-olds and adults were not significant for both variables. The results of the Scheffé post hoc tests are displayed in Table 5.

To test whether successful inhibition (i.e., the number of delayed intrusion errors) had an influence on inference generation in any age group, a two-factorial ANOVA was calculated for every age group, with the three types of reaction times (C, I-SV, I-OV) as within-subject factor, and delayed intrusion errors as covariate. A significant interaction between reaction times and delayed intrusion errors was found in the 8-year-olds' group, $F(2, 60) = 3.16$, $p = .050$, $\eta_p^2 = .10$, whereas no significant interaction existed for the 10-year-olds and adults (both $F < 1$).

In order to examine the influence of inhibition on the 8-year-olds' reaction times more closely, we divided the 8-year-old participants into two groups with high or low inhibition using the median of delayed intrusion errors (7) as a cut-off point. Two *t*-tests for the two types of inferences were conducted for each group. The group with relatively many intrusion errors did

not have significant reaction-time differences: exact inferences, $t(15) = -2.09$, $p = .054$; valence inferences, $t(15) = -1.36$, $p = .193$. The group with relatively few intrusion errors had significantly shorter reaction times in the congruent condition compared to I-OV: $t(15) = -3.90$, $p = .001$. The difference between the congruent condition and I-SV was not significant, $t(15) = -1.09$, $p = .292$. This result is shown in Figure 2. Our assumption that good inhibition is important for children's inference generation was partly supported for the 8-year-olds, but not for the 10-year-olds.

– Enter Figure 2 about here –

Discussion

The aim of our study was to identify factors that influence children's and adult's ability to generate emotional inferences. Unlike other studies concerning children's inference processes, we used an online method in order to assess the processes that happen *during* as opposed to *after* reading. As possible influencing factors, we focused on two features of the texts and one variable of the participants. First, we alternated whether the text was presented aurally or whether the participants read the texts. Second, we presented the texts in two different versions, varying whether or not the text elaborated on the protagonist's goal. Third, as a variable of the participants, we assessed their working memory's ability to update information, especially to suppress information that is no longer relevant. The results reveal the influences of all three factors, and the age of the participants moderated the influence of every factor.

Analysis of the descriptive statistics showed that the error rates were very low. This demonstrates that the reaction-time task was comprehensible and feasible for every age group. Reaction times got faster with age, as expected in this sort of test. Similarly, the mean reading times for the stories in the reading condition were shorter for older children and adults. The scores on the comprehension questions were high, which allows us to assume that the texts were comprehendible for all age groups. The scores also got higher with age.

One research question in the current study was whether children and adults build online emotional inferences while reading or listening to texts. We operationalized emotional inferences by comparing reaction times for reactions to target sentences mentioning emotions that were congruent to the protagonist's emotion in the story with reaction times to emotions that were incongruent. If the reader made an inference about the protagonist's emotion, this emotion should be easily accessible and therefore allow a quicker reaction than an incongruent emotion. By calculating a multifactorial ANOVA, we analyzed not only whether emotional inferences were built but also whether this was influenced by the type of presentation (reading vs. listening) or by goal emphasis.

The 8-year-olds' results were not influenced by medium or goal emphasis, so our hypothesis about the positive influence of auding compared to listening and the positive influence of goal emphasis could not be confirmed. However, our assumption concerning emotional inferences was confirmed, in that 8-year-olds built emotional inferences of emotional valence, but no exact emotional inferences. This replicated the findings of both our previous studies (Diergarten & Nieding, 2015), in which 8-year-old children also only built emotional valence inferences.

Both the 10-year-olds' and the adults' emotional inferences were influenced by the type of text presentation, in that both age groups seem to have profited from the reading condition. For the read texts, the reaction times for congruent target sentences were significantly shorter compared to reaction times for incongruent target sentences. In both age groups, we found only exact emotional inferences, but no inferences of emotional valence (we will come back to this point later in the discussion). In the listening condition, we did not find emotional inferences. These results confirmed our hypotheses only concerning the adults. For the two groups of children, we expected the auding condition to support emotional inferences. This hypothesis was based on the findings of Stitch and James (1984) that reading performance doesn't reach the level of auding performance until seventh grade. Our results, however, suggest that, at least

for building emotional inferences, reading is superior to auding even for children in fourth grade, and is comparable to auding for second-graders. We have two sets of explanations for these findings.

First, the results could be due to the fact that our participants were native German speakers. A meta-analysis by Florit and Cain (2011) revealed that the influence of decoding on reading comprehension decreases and the influence of listening comprehension increases at an earlier stage for readers of transparent orthographies, such as German orthography. English, on the other hand, has a complex orthography. Studies with English samples have found that reading comprehension was still strongly related to decoding ability even after three to five years of reading instruction with less influence on listening comprehension (Florit & Cain, 2011). A German replication of the simple view of reading studies, however, found that reading comprehension was already highly influenced by listening comprehension in the middle of second grade (mean age of student 8;1; see Marx & Jungmann, 2000). This could explain why our 8-year-olds performed equally well in the reading condition.

Another explanation lies in the fact that we had our participants read the texts out loud and not silently, as is usually the case in studies comparing listening and reading comprehension. Reading research comparing silent and oral reading has revealed that the age of the reader determines whether silent or oral reading is beneficial. A study by Prior et al. (2011) suggests that elementary school children profit from reading out loud up until fifth grade, while silent reading is either equivalent (Franklin, Mooneyham, Baird, & Schooler, 2014) or better (Prior et al., 2011) for comprehension than reading out loud for older students and adults. Prior et al. explain this age effect by arguing that reading aloud helps younger and less-skilled readers to keep their attention high and that it prevents them from skipping over difficult words. Reading aloud also allows them to process the information with two senses, that is, seeing the text and at the same time listening to their own voice. We assume that the fourth-graders

profited from reading out loud and therefore the reading condition was more helpful overall for their text comprehension and emotional inferences.

That adults built emotional inferences only in the reading and not in the listening condition confirmed our hypothesis. The positive effect of reading compared to auding in adults is in agreement with the literature (e.g., Kürschner et al., 2006; Rubin et al., 2000). An explanation for this common finding is the amount of mental effort that is invested in different media. Zumbach and Schwartz (2014) found that adults invested more mental effort in reading than listening, resulting in higher knowledge acquisition in the reading condition. They argue that participants judged reading to be harder compared to listening and therefore invested more mental effort while reading. Zumbach and Schwartz (2014) theoretically support this assumption by referring to Geary's (2004) differentiation between biologically primary and secondary knowledge. According to Geary's theory, processing auditory information is a primary skill, while reading is a secondary skill. For this reason, listening appears to be easier than reading, so that less mental effort is invested, although elaborative listening would require the same amount of concentration (Rubin et al., 2000).

It is noteworthy, however, that reading was superior to listening in spite of the fact that our participants had to read out loud, which, for adults, is a disadvantage compared to silent reading. Once a person masters silent reading, oral reading becomes distracting, taking attention away from the content to focus on pronunciation and articulation (Prior et al., 2011; Prior & Welling, 2001). Apparently, the advantages of reading compared to listening were still greater in spite of having to read out loud. However, one limitation of this study is that reading aloud vs. silently and the difference between reading and listening were confounded. It is therefore not possible to make any conclusions about emotional inference during silent reading based on our data.

The mental effort theory is also a possible explanation for the children's better results in the reading condition. Many children in Germany listen to audio plays frequently

(Medienpädagogischer Forschungsverbund Südwest, 2013). Audio plays are very often used secondarily, that is, running in the background while the child is playing, drawing, or busy with another activity (Paus-Haase, Hölterschinken, & Tietze, 1990). For this reason, children might be used to not paying close attention to the text when listening to an audio play, while they know from school that it is important to pay attention to the content of written texts.

The lack of evidence of valence inferences for the 10-year-olds and adults seems puzzling at first because building exact inferences includes a representation of the emotional valence. This replicates a finding from our previous study (Diergarten & Nieding, 2015), in which the same pattern was found in the group of adults. We explained it with the "novel-popout" effect (Johnston & Hawley, 1994, p. 56). This effect describes the phenomenon that novel cues are easier to detect among distractors than familiar cues. The older participants integrated the protagonist's emotion into their situation model, thereby familiarizing with it. An emotion with the opposite valence appeared so different from the familiar emotion that it popped out, facilitating reaction times, whereas an emotion of the same valence would be familiar enough not to cause a popout effect. This explanation is in accordance with Treisman and Gelade's (1980) feature-integration theory, which states that processing times are longer when targets and distractors are similar.

Another research question concerned the influence of goals in stories on emotional inferences. Adults built emotional inferences in the goal-emphasis condition but not in the no-goal-emphasis condition. This shows that they clearly profited from goal emphasis. The children, on the other hand, didn't seem to profit from the goal-emphasis condition, as there was no interaction between reaction times and the goal-emphasis condition. While the results of the adult participants fit well with previous research demonstrating a positive influence of goals on story comprehension (Egidi & Gerrig, 2006; Lynch & van den Broek, 2007), the children's data contradict our assumption that goals should be especially helpful for children's story comprehension. A possible explanation for this finding could be the way in which we

implemented goal emphasis. The goal was emphasized by two sentences that explained why the goal was important to the protagonist. In the example given in the method section, the sentences explained that Thomas was doing a report about lions for school. By stating the goal this way, a new set of information was introduced that needed to be incorporated into the situation model. In this example, participants might have built a mental representation of Thomas receiving this assignment from his teacher, which would mean a shift in the time and space dimension of the situation model. These shifts are known to require an updating of the situation model (Radvansky & Copeland, 2010). Another updating was necessary when the story shifted back to the main events after those two sentences. In the no-goal-emphasis condition, the two additional sentences did not involve such dimensional shifts. In the story about Thomas, for example, the sentences described him visiting other animals in the zoo, thus remaining in the same time and space dimensions. Thus, the goal-emphasis condition might have tied up additional mental capacity in the required updating process, which might have annulled the positive effects that a goal elaboration might otherwise have had. Further research is needed, however, to examine if this explanation is true. This could be achieved by adding reaction-time tasks after the critical sentences or measuring the reading times for those sentences (Radvansky & Copeland, 2001, 2010). Also, further research should again address the question of whether goals play a role in children's emotional inference by realizing a simpler goal-emphasis version, for example, by only stating "this is very important for . . ." without explaining the reason why it is so important.

Concerning updating, we found a significant effect in the group of 8-year-olds: children with relatively few intrusion errors built emotional inferences (valence), while children with many intrusion errors did not. This finding is in agreement with previous studies that found that working memory is important for text comprehension (Cain et al., 2004; Oakhill et al., 2005), and that the updating function of working memory, in particular, is relevant for comprehension at the situation-model level (Radvansky & Copeland, 2001). This is understandable considering

that maintaining a coherent situation model requires constant renewal of the information stored in the working memory, replacing information that is no longer relevant with new information. In our study, this effect only occurred in the group of 8-year-olds. This is probably due to the fact that we developed the stories to be adequate for the reading level of 8-year-olds. For this reason, the stories were demanding enough for 8-year-olds, and thus building a profound situation model including emotional inferences required a high updating capacity from them. However, 10-year-olds and adults naturally had a higher average in updating capacity; also, the stories were very easy for them to comprehend, so that individual differences in updating capacity were not relevant in these age groups.

Limitations and Conclusions

When designing the stories for our experiment, we were challenged by the great age difference between our participants. We wanted to include adults in order to maintain comparability with other research concerning emotional inferences, as there has been hardly any work on emotional inferences in children. One option would have been to construct separate stories for every age group, so that they would be challenging enough for everybody. However, this procedure would have made it difficult to compare the results between age groups, which was the reason why we chose to use the same stimuli for all. Future research could try to replicate the findings with age-adjusted material for all age groups. Such research should also include one or two additional age groups between 10- year-olds and adults (for instance, middle school or early high school students) in order to answer questions about developmental changes in this age range.

Additionally, the interpretation of our results concerning differences between reading and listening is limited by the fact that our reading condition was actually a reading-aloud condition. This is problematic because the effects of reading aloud compared to reading silently are subject to developmental changes, so that our results might be disturbed by an interaction

effect and therefore could not be clearly attributed to the effects of reading vs. listening. Further research should include an additional silent-reading condition to address this problem.

Despite some limitations, our study made an important contribution to the field of situation model and inference research. By using an online method, we assessed cognitive processes that happen during reading and that therefore approximate natural comprehension processes. Considering that our previous studies (Diergarten & Nieding, 2015) also demonstrated online emotional inferences, the current study constitutes confirmation that the applied online measure for emotional inferences works. At the same time, the current data complement our prior data regarding emotional inferences in movies and audio plays by adding information about those processes during reading and listening to short stories.

Our results are encouraging concerning children's ability to learn about emotional knowledge. As Nelson (2007) pointed out, narratives are an important information source for children to learn about other people's inner worlds, as they gain insights about how different individuals feel and experience themselves and others. Indeed, several training studies have revealed that listening to narratives has positive effects on emotional knowledge and theory of mind development (Gavazzi & Ornaghi, 2011; Guajardo & Watson, 2002; Ornaghi, Brockmeier, & Gavazzi, 2011). Although the named studies also showed that the effects are higher if children are engaged in discussions about emotions or mental states after reading, the present study asserts that children think about the protagonist's emotional state even when they are not explicitly asked to do so, namely, during natural, private reading sessions.

References

Ackerman, B. P. (1984). Storage and processing constraints on integrating story information in children and adults. *Journal of Experimental Child Psychology, 38*(1), 64–92. doi:10.1016/0022-0965(84)90019-5

Ackerman, B. P., & McGraw, M. (1991). Constraints on the causal inferences of children and adults in comprehending stories. *Journal of Experimental Child Psychology, 51*(3), 364–394. doi:10.1016/0022-0965(91)90083-5

Brown, D. P. (1954). *Auding as the primary language ability* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671–681. doi:10.1037/0022-0663.96.4.671

Carretti, B., Cornoldi, C., Beni, R. de, & Romano, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology, 91*(1), 45–66. doi:10.1016/j.jecp.2005.01.005

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., . . . Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132*(1), 113–132. doi:10.1037/0096-3445.132.1.113

Diergarten, A. K. & Nieding, G. (2015). Children's and adults' ability to build online emotional inferences during comprehension of audiovisual and auditory texts. *Journal of Cognition and Development, 16,* 381-406. doi: 10.1080/15248372.2013.848871

Egidi, G., & Gerrig, R. J. (2006). Readers' experiences of characters' goals and actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1322–1329. doi:10.1037/0278-7393.32.6.1322

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, *23*(4), 553–576. doi:10.1007/s10648-011-9175-6

Franklin, M. S., Mooneyham, B. W., Baird, B., & Schooler, J. W. (2014). Thinking one thing, saying another: The behavioral correlates of mind-wandering while reading aloud. *Psychonomic Bulletin & Review*, *21*(1), 205–210. doi:10.3758/s13423-013-0468-2

Gavazzi, I. G., & Ornaghi, V. (2011). Emotional state talk and emotion understanding: A training study with preschool children. *Journal of Child Language*, *38*(5), 1124–1139. doi:10.1017/S0305000910000772

Geary, D. C. (2004). Evolution and cognitive development. In R. L. Burgess & K. MacDonald (Eds.), *Evolutionary perspectives on human development* (pp. 99–133). Thousand Oaks, CA: Sage Publications.

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. (1992). Do readers mentally represent characters' emotional states? *Cognition & Emotion*, *6*(2), 89–111. doi:10.1080/02699939208411061

Gernsbacher, M. A., & Robertson, R. R. (1992). Knowledge activation versus sentence mapping when representing fictional characters' emotional states. *Language and Cognitive Processes*, *7*(3-4), 353–371. doi:10.1080/01690969208409391

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *RASE: Remedial & Special Education*, *7*(1), 6–10. doi:10.1177/074193258600700104

Graesser, A. C., Bertus, E. L., & Magliano, J. P. (1995). Inference generation during the comprehension of narrative text. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of*

coherence in reading (pp. 295–320). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Guajardo, N. R., & Watson, A. C. (2002). Narrative discourse and theory of mind development. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *163*(3), 305–325. doi:10.1080/00221320209598686

Gygax, P., Oakhill, J., & Garnham, A. (2003). The representation of characters' emotional responses: Do readers infer specific emotions? *Cognition & Emotion*, *17*(3), 413–428. doi:10.1080/02699930244000048

Hildyard, A., & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes*, *1*, 91–117. doi:10.1080/01638537809544431

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. doi:10.1007/BF00401799

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

Johnston, W. A., & Hawley, K. J. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. *Psychonomic Bulletin & Review*, *1*(1), 56–72. doi:10.3758/BF03200761

Kendeou, P., Bohn-Gettler, C., White, M. J., & van den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259–272. doi:10.1111/j.1467-9817.2008.00370.x

Kendeou, P., Lynch, J. S., van den Broek, P., Espin, C. A., White, M. J., & Kremer, K. E. (2005). Developing successful readers: Building early comprehension skills through television viewing and listening. *Early Childhood Education Journal*, *33*(2), 91–98. doi:10.1007/s10643-005-0030-6

Kürschner, C., Schnotz, W., & Eid, M. (2006). Konstruktion mentaler Repräsentationen beim Hör- und Leseverstehen [The construction of mental representations in listening and

reading comprehension]. *Zeitschrift für Medienpsychologie*, *18*(2), 48–59. doi:10.1026/1617-6383.18.2.48

Kürschner, C., Schnotz, W., Eid, M., & Hauck, G. (2005). Individuelle Modalitätspräferenzen beim Textverstehen: Präferenzen für auditive oder visuelle Sprachverarbeitung in unterschiedlichen Bevölkerungsgruppen [Individual modality preferences in text comprehension: Preferences for auditory or visual language processing in different populations]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *37*(1), 2–16. doi:10.1026/0049-8637.37.1.2

Lynch, J. S., & van den Broek, P. (2007). Understanding the glue of narrative structure: Children's on- and off-line inferences about characters' goals. *Cognitive Development*, *22*(3), 323–340. doi:10.1016/j.cogdev.2007.02.002

Marx, H., & Jungmann, T. (2000). Abhängigkeit der Entwicklung des Leseverstehens von Hörverstehen und grundlegenden Lesefertigkeiten im Grundschulalter: Eine Prüfung des Simple View of Reading-Ansatzes [Dependency of reading comprehension development on listening and basic reading skills]. *Zeitschrift fur Entwicklungspsychologie und Pädagogische Psychologie*, *32*(2), 81–93. doi:10.1026//0049-8637.32.2.81

Medienpädagogischer Forschungsverbund Südwest (mpfs, 2013). (2013). *KIM 2012 - Kinder + Medien, Computer + Internet. Basisstudie zum Medienumgang sechs-bis 13-Jähriger in Deutschland* [KIM 2012 - Children + media, computer + internet. Basic study about six- to 13-year-olds' media use in Germany.]. Stuttgart. Retrieved from Medienpädagogischer Forschungsverbund Südwest (mpfs) website: http://www.mpfs.de/fileadmin/KIM-pdf12/KIM_2012.pdf

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe'

tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. doi:10.1006/cogp.1999.0734

Molinari, C., Burin, D., Saux, G., Barreyro, J. P., Irrazabal, N., Bechis, M. S., . . . Ramenzoni, V. (2009). Fictional characters' emotional state representation: What is its degree of specificity? *Psicothema*, *21*(1), 9–14.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*(2), 111–121. doi:10.1111/j.2044-8295.1990.tb02349.x

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory*. Cambridge, Mass.: Harvard University Press.

Nieding, G., & Ohler, P. (2008). Mediennutzung und Medienwirkung bei Kindern und Jugendliche [Media use and media effects in children and adolescents]. In B. Batinic (Ed.), *Medienpsychologie* (pp. 379–400). Heidelberg: Springer.

Oakhill, J. V., & Cain, K. (2004). The development of comprehension skills. In T. Nunes & P. Bryant (Eds.), *Handbook of Children´s Literacy* (pp. 155–189). Dordrecht: Kluwer Academic Publishers.

Oakhill, J. V., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, *18*(7-9), 657–686. doi:10.1007/s11145-005-3355-z

Ohler, P., Nieding, G., & Töpper, J. (2002). Die Repräsentation von Protagonistenemotionen bei der Enkodierung narrativer Filme [The representation of the protagonist’s emotions during the encoding of narrative films]. In M. Baumann, A. Keinath, & J. F. Krems (Eds.), *Experimentelle Psychologie* (p. 70). Regensburg: Roderer.

Ornaghi, V., Brockmeier, J., & Gavazzi, I. G. (2011). The role of language games in children's understanding of mental states: A training study. *Journal of Cognition and Development, 12*(2), 239–259. doi:10.1080/15248372.2011.563487

Orrantia, J., Múñez, D., & Tarín, J. (2014). Connecting goals and actions during reading: the role of illustrations. *Reading and Writing, 27*(1), 153–170. doi:10.1007/s11145-013-9437-4

Palladino, P., Cornoldi, C., Beni, R. de, & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition, 29*(2), 344–354. doi:10.3758/BF03194929

Paus-Haase, I., Hölterschinken, D., & Tietze, W. (1990). *Alte und neue Medien im Alltag von jungen Kindern. Orientierungshilfen für Eltern und Erzieher* [Old and new media in young children's daily routines]. Freiburg: Lambertus-Verlag.

Prior, S. M., Fenwick, K. D., Saunders, K. S., Ouellette, R., O'Quinn, C., & Harvey, S. (2011). Comprehension after oral and silent Reading: Does grade level matter? *Literacy Research and Instruction*, *50*(3), 183–194. doi:10.1080/19388071.2010.497202

Prior, S. M., & Welling, K. A. (2001). 'Read in your head': A Vygotskian analysis of the transition from oral to silent reading. *Reading Psychology*, *22*(1), 1–15. doi:10.1080/02702710151130172

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. *Memory & Cognition*, *29*(8), 1073–1080. doi:10.3758/BF03206375

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American Journal of Psychology, 117*(2), 191–213. doi:10.2307/4149022

Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 210–216. doi:10.1037/a0017258

Radvansky, G. A., Gibson, B. S., & McNerney, M. W. (2014). Working memory, situation models, and synesthesia. *The American Journal of Psychology*, *127*(3), 325–342. doi:10.5406/amerjpsyc.127.3.0325

Rey, G. D. (2012). *Methoden der Entwicklungspsychologie: Datenerhebung und Datenauswertung* [Methods of developmental psychology. Data collection and data analysis]. Norderstedt bei Hamburg: BoD.

Rickheit, G., & Strohner, H. (1983). Medienspezifische Textverarbeitung [Media specific processing of texts]. In P. Finke (Ed.), *Sprache im politischen Kontext [Language in a political context]* (pp. 175–207). Tübingen: Max Niemeyer Verlag.

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & Cognition, 31*(8), 1284–1292. doi:10.3758/BF03195811

Roseman, I. J., Antoniou, A. A., & Jose, P. E. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, *10*(3), 241–278. doi:10.1080/026999396380240

Rubin, D. L., Hafer, T., & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education*, *49*(2), 121–133. doi:10.1080/03634520009379200

Stitch, T. G., & James, J. H. (1984). Listening and reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 293–317). New York: Longman.

Therriault, D. J., & Rinck, M. (2007). Multidimensional situation models. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 311–327). Mahwah, NJ: Lawrence Erlbaum Associates.

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading and Writing, 26*(3), 403–429. doi:10.1037/t15094-000

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136. doi:10.1016/0010-0285(80)90005-5

Unsöld, I. H. (2008). *Die Bildung von Inferenzen bei der kognitiven Verarbeitung medialer Texte* [The formation of inferences during processing of media texts: A study with children and adults]. Hamburg: Dr. Kovač.

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

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258. doi:10.1111/j.1467-9922.2010.00593.x

Zumbach, J., & Schwartz, N. (2014). Hyperaudio learning for non-linear auditory knowledge acquisition. *Computers in Human Behavior, 41*(0), 365–373. doi:10.1016/j.chb.2014.10.024

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*(5), 292–297. doi:10.1111/j.1467-9280.1995.tb00513.x

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 386–397. doi:10.1037/0278-7393.21.2.386

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185. doi:10.1037/0033-2909.123.2.162

Table 1

Example story in goal-emphasis and no goal-emphasis condition and target sentences

With goal-emphasis:
Today is the first gym class after summer break. Sandra and her class have long jump training. The teacher measures the distances. It's Sandra's turn now. *Sandra's long jump performance didn't use to be so good. But in the summer break, she did a lot of training. The best kids may join the school team and take part in a big long jump competition.* After training, the teacher compliments Sandra on her result. He says that Sandra achieved the best jumping distance of the whole class.

Without goal-emphasis:
Today is the first gym class after summer break. Sandra and her class have long jump training. The teacher measures the distances. It's Sandra's turn now. *She bends down to tie her shoes firmly so the laces won't come undone during takeoff. Then she sprints to the pit, takes off and lands at a far distance in the pit.* After training, the teacher compliments Sandra on her result. He says that Sandra achieved the best jumping distance of the whole class.

Target emotion: Pride.

Target sentences:

Congruent (C): Sandra feels proud.

Incongruent – Same Valence (I-SV): Sandra feels happy.

Incongruent – Opposite Valence (I-OV): Sandra feels angry / Sandra feels sad.

Note. Stories were presented in German. This is a translation.

Table 2

Means and SDs for errors, reaction times, reading times, and comprehension questions

	8-year-olds	10-year-olds	Adults
Percentage of errors in reaction-time task	1.37% (3.07)	1.17% (2.94)	0.78% (1.10)
Mean reaction times (in milliseconds)	1084 (139)	918 (135)	723 (120)
Average reading time for one story (in seconds)	62.19 (16.53)	41.68 (9.67)	24.92 (2.28)
Comprehension questions (Max = 16)	14.80 (1.10)	15.59 (.69)	15.80 (.44)

Table 3

Mean reaction times and SDs in the three congruency conditions and results of the t*-tests testing inference generation in the reading and listening condition*

		8-year-olds		**10-year-olds**		**Adults**	
Valence inferences		**C**	**I-OV**	**C**	**I-OV**	**C**	**I-OV**
	Reading	1019 (187)	1105 (165)	883 (159)	932 (174)	695 (158)	701 (107)
		$t = -3.03$**		$t = -1.90$		$t = -0.29$	
	Listening	1029 (151)	1118 (169)	916 (165)	934 (141)	760 (143)	731 (133)
		$t = -2.65$*		$t = -0.79$		$t = 1.61$	
Exact inferences		**C**	**I-SV**	**C**	**I-SV**	**C**	**I-SV**
	Reading	1019 (187)	1098 (182)	883 (159)	940 (158)	695 (158)	728 (141)
		$t = -2.45$*		$t = -2.67$*****		$t = -2.09$*****	
	Listening	1029 (151)	1075 (197)	916 (165)	873 (139)	760 (143)	718 (122)
		$t = -1.54$		$t = 1.73$		$t = 1.87$	

Note. If the difference between C and I-OV is significant and the *t*-value positive, this indicates inferences of the emotional valence, while a significant difference between C and I-SV with a positive *t*-value indicates exact emotional inferences. The number of degrees of freedom (df) is 31 in all tests.
*$p < .05$ **$p < .01$

Table 4

Mean reaction times and SDs in the three congruency conditions and results of the t*-tests testing inference generation in the goal-directed and non-goal-directed condition*

		8-year-olds		**10-year-olds**		**Adults**	
Valence inferences		**C**	**I-OV**	**C**	**I-OV**	**C**	**I-OV**
	Goal-emphasis	1016 (165)	1102 (159)	875 (154)	939 (175)	692 (116)	716 (118)
		$t = -2.84$**		$t = -2.55$*		$t = -1.96$	
	No goal-emphasis	1032 (175)	1121 (182)	923 (143)	928 (152)	764 (168)	716 (115)
		$t = -2.66$*		$t = -0.20$		$t = 2.40$****	
Exact inferences		**C**	**I-SV**	**C**	**I-SV**	**C**	**I-SV**
	Goal-emphasis	1016 (165)	1074 (190)	875 (154)	919 (122)	692 (116)	728 (139)
		$t = -1.64$		$t = -2.20$****		$t = -2.25$*	
	No goal-emphasis	1032 (175)	1099 (207)	923 (143)	894 (172)	764 (168)	716 (113)
		$t = -1.74$		$t = 1.10$		$t = 2.06$*	

Note. If the difference between C and I-OV is significant and the *t*-value positive, this indicates inferences of the emotional valence, while a significant difference between C and I-SV with a positive *t*-value indicates exact emotional inferences. The number of degrees of freedom (df) is 31 in all tests.
*$p < .05$ **$p < .01$

Table 5

Mean numbers and SDs of correctly recalled items and delayed intrusion errors in the three age groups as well as the p*-values of the Scheffé post hoc tests*

	8-year-olds	**10-year-olds**	**Adults**
Correct recall	22.06 (3.53)	24.38 (3.64)	26.13 (3.17)
Scheffé post hoc tests	$p = .032$		
		$p = .134$	
	Between 8-year-olds and adults: $p < .001$		
Delayed intrusion errors	6.69 (3.19)	4.75 (2.79)	3.38 (2.21)
Scheffé post hoc tests	$p = .023$		
		$p = .143$	
	Between 8-year-olds and adults: $p < .001$		

Note. Max = 30 for both variables

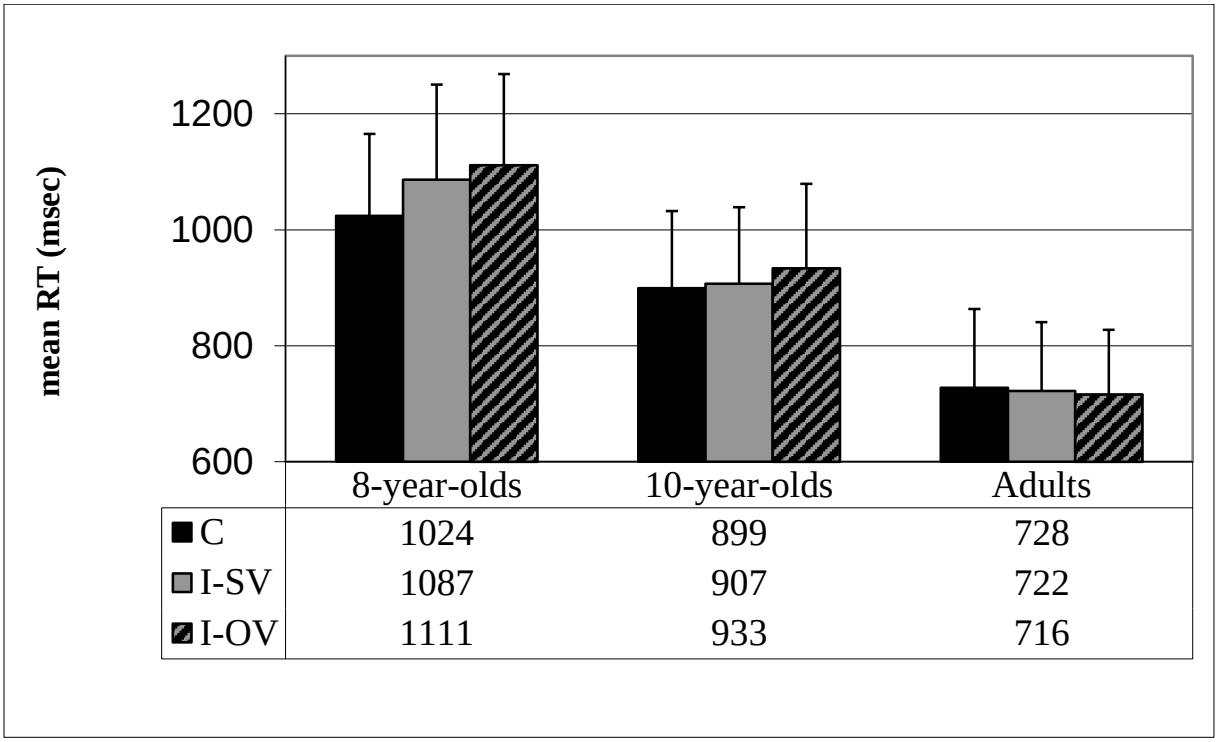

Figure 1. Reaction times in the three congruency conditions.

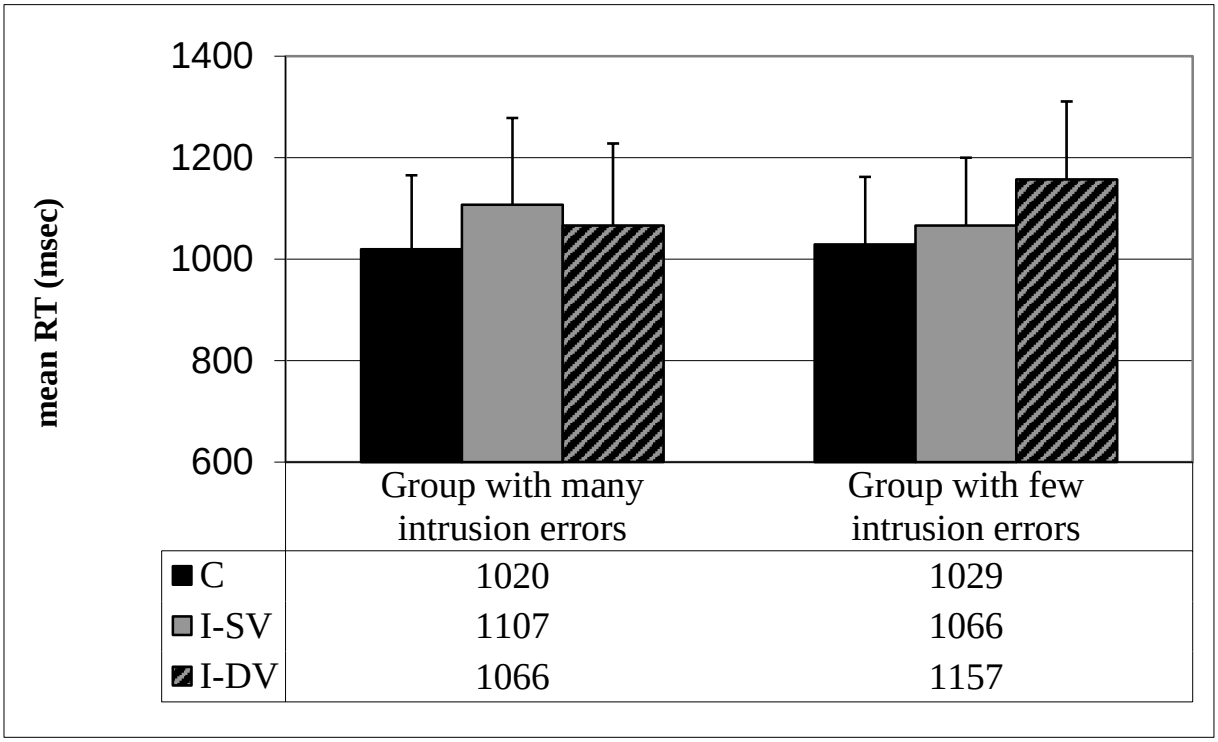

Figure 2. The 8-year-olds' reaction times in the three congruency conditions, divided into two groups with many vs. few intrusion errors in the updating task.