DISCOURSE PROCESSES, *40*(3), 239–257

Perspective Effects on Expository Text Comprehension: Evidence From Think-Aloud Protocols, Eyetracking, and Recall

Johanna K. Kaakinen and Jukka Hyönä
University of Turku, Finland

In this study, 36 participants read an expository text describing 4 rare illnesses from a given perspective. Their eye movements were recorded during reading, and think-alouds were probed after 10 relevant and 10 irrelevant sentences. A free recall was collected after reading. The results showed that in addition to increasing the fixation time and recall for relevant in comparison to irrelevant text information, a reading perspective guides readers to use slightly different comprehension processes for relevant text information, as shown by think-aloud protocols. Repetitions were more frequent responses after relevant than after irrelevant target sentences. Verbally reported processing strategies were associated with the eye-fixation patterns. Verbal responses indicative of deeper processing were associated with longer first-pass fixation times than those indicative of shallower processing. It is concluded that a "triangulation" using complementary measures is a worthwhile endeavor when studying text-comprehension processes.

Let us imagine that a friend has fallen ill with a disease we have never heard of, trigeminusneuralgy. We find a journal article, which covers some rare diseases, and we read it to learn about our friend's disease. When we are orientated like this, it can be said that we have a specific comprehension perspective from which we approach the article. The aim of this study was to examine how a reading perspective, like the one described, influences the processing of expository text. The newsworthy aspect of this study is that online text comprehension was studied by simultaneously combining two methodologies, eyetracking and think-alouds. This

Correspondence should be addressed to Johanna Kaakinen, Florida State University, Department of Psychology, Tallahassee, FL 32306–1270. E-mail: johanna.kaakinen@utu.fi

allows us to examine the relation between these two processing measures. More specifically, the quality of processing, as revealed by think-aloud protocols, could be directly related to the processing time measures. To our knowledge, this study is the first one in which online text comprehension processes were simultaneously studied with eye-tracking and think-aloud methodology.

A reading perspective defines what information in the text is relevant and what information is irrelevant for the reader (e.g., Pichert & Anderson, 1977). When reading the article describing the rare illnesses, we would find only information about trigeminusneuralgy interesting and regard information concerning other diseases relatively useless for our purposes. After reading the article, we remember more information relevant to our reading perspective than irrelevant to our reading perspective, even if we changed the perspective at recall and tried to recall new details (e.g., Anderson, Pichert, & Shirey, 1983; Baillet & Keenan, 1986; Kaakinen, Hyönä, & Keenan, 2001; but see Anderson & Pichert, 1978). It appears that the perspective active during reading guides the reader's attentional resources so that the perspective-relevant text information is processed more carefully (i.e., slower) than perspective-irrelevant text information (Goetz, Schallert, Reynolds, & Radin, 1983; Rothkopf & Billington, 1979), especially if the reader does not have prior knowledge related to the text contents (Kaakinen, Hyönä, & Keenan, 2002, 2003). For example, in our study (Kaakinen et al., 2003) participants read an expository text describing rare diseases of which they had no prior knowledge. A reading perspective was induced before reading by instructing the participants to imagine that a friend has fallen ill with one of the diseases described in the text. Readers' eye movements were recorded during reading, and a free recall was collected afterward. The readers showed longer eye fixation times on perspective-relevant than on perspective-irrelevant sentences during the initial reading of the target sentences (i.e., the first-pass reading). Moreover, they spent more time rereading the perspective-relevant than the perspective-irrelevant sentences. After reading, the participants had superior memory for perspective-relevant in comparison to perspective-irrelevant information.

The question remains: What types of processes are carried out during the extra time devoted to relevant text information to yield better memory for relevant than for irrelevant text information? Why do readers spend more time on perspective-relevant information? According to the *standards of coherence framework*, a reading goal defines the standards of coherence the reader maintains during reading (van den Broek, Lorch, Linderholm, & Gustafson, 2001; van den Broek, Risden, & Husebye-Hartmann, 1995; see also Goldman, Varma, & Coté, 1996). The standards of coherence dictate the criteria for sufficient understanding, and the reader adjusts the comprehension processes to meet these criteria. For example, in the study of van den Broek et al. (2001), participants read an expository text either for the purpose of study or for pleasure. Readers were asked to think-aloud while reading. The analysis of the think-aloud protocols revealed that when reading for study, readers made more coherence-enhancing inferences, such as explaining and predicting, than when reading for pleasure. Moreover, they frequently repeated and paraphrased text. When reading for pleasure, on the other hand, readers were not that much concerned about the coherence of the text but tried to relate the text contents to their personal experiences by making associative inferences and evaluations of the text contents. The subsequent recall performance was better in the study condition than in the entertainment condition.

It may be that when the reader approaches the text from a specific perspective, the criterion for comprehension is to obtain a coherent memory representation (see Coté, Goldman, & Saul, 1998) for the perspective-relevant information in the text, whereas the perspective-irrelevant text information may be processed more superficially. Readers construct different levels of memory representations for the text: *a surface-level representation*, *a textbase representation,* and *a situation model* (Kintsch, 1998; van Dijk & Kintsch, 1983). The surface code preserves the particular words and clauses presented in the text and the linguistic relations between them. The surface code is part of the textbase, which is constructed of the core meanings of the clauses—propositions—and the connections between the propositions. Thus, the textbase consists of propositions and connections of these meaning elements presented in the text. However, simply retaining the text propositions and the order in which they are presented in the text does not necessarily mean that the reader has understood what the text is about (e.g., Moravcsik & Kintsch, 1995). Sometimes it is necessary for the reader to employ prior knowledge (linguistic knowledge, world knowledge, knowledge about the communicative situation, or personal experience, see Kintsch, 1998) to obtain sufficient understanding of the text. The situation model corresponds to a deeper level of understanding of the text, and it is constructed by elaborating and integrating text information with prior knowledge. Previous research on perspective effects on text memory suggests that readers construct a better textbase for the relevant than for the irrelevant text information (e.g., Anderson et al., 1983; Baillet & Keenan, 1986; Kaakinen et al., 2001). For example, Baillet and Keenan showed that readers produce in a free-recall task more perspective-relevant than perspective-irrelevant information and that they are not able to retrieve originally irrelevant text information even if they were given a new perspective as a recall cue. This constraining effect of a reading perspective on text recall is evident soon after reading and even more so after a 1-week delay. Thus, it can be hypothesized that readers engage in textbase construction activities such as repeating parts of text and paraphrasing the contents of text segments when they encounter relevant information in text.

However, do readers also engage in more rigorous situation model building for relevant than for irrelevant information during the course of reading? Lehman and Schraw (2002) used different offline comprehension measures to examine how a reading perspective affects textbase and situation model construction. Participants read a text either with or without a perspective, and the memory for text was tested with a recognition test, a recall task, and an essay task in which the participants had to explain why certain events described in the text had occurred. Readers in the perspective condition produced essays that reflected deeper processing (i.e., more inferences and better integration of these inferences to the text and to each other) than the readers in the no-perspective condition. However, because Lehman and Schraw did not include a perspective shift condition in their experimental design, it is difficult to say whether the reading perspective constrains the situation model building during the course of reading or whether the perspective effect appears only at recall. If readers do attempt to build a more comprehensive situation model for the relevant than for the irrelevant text information, we should observe an increase in situation model construction activities such as self-explanations or predictions when readers encounter relevant information in the text.

The aim of this study was to examine the online comprehension of relevant and irrelevant text information by combining eye-tracking measures (see, e.g., Hyönä, Lorch, & Rinck, 2003) and think-aloud protocols (see, e.g., Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). Combining different measures allows us to examine the validity of the verbal reports. So far, only a few attempts are made to relate them to objective performance (for the need of triangulation of online and offline measures, see also Graesser & Magliano, 1991). Olson, Duffy, and Mack (1984) had a group of readers provide think-alouds online while they read well-formed and ill-formed stories. Another group of readers read the same texts for comprehension while their sentence-reading times were collected. Regression analyses revealed that those sentences in the well-formed stories that were more likely to produce inferences and predictions during the think-aloud procedure were associated with relatively longer reading times. Such a relation did not exist for the ill-formed texts. Fletcher (1986) and Trabasso and Suh (1993) found a modest but significant correlation between the sentence-reading time and the likelihood that the sentence contains information mentioned previously in the think-aloud protocols. A sentence was found to be easier to read when it was closely linked to the contents of the evolving mental representation of the text.

Even though these results suggest that reported inferential activities are related to differential reading times, the verbal reports and reading times have not been collected simultaneously. On the other hand, Coté and Goldman (1999) and Coté et al. (1998) did collect parallel sentence-reading times and think-aloud protocols, but they did not report the reading time data because in their study participants read the texts aloud. Reading aloud caused uncontrolled variability in the recorded reading times, namely individual differences in the speaking pace. Moreover, because their participants were children, the experimenter had to occasionally prompt to help in reading, causing inaccuracies in the reading time measurements. The exact nature of the relation between the verbal responses and the online processing is thus still unknown. For example, is the mental effort devoted for deeper understanding reflected in the initial reading of the sentences, or does it only increase rereading of the sentences? In this study, readers read the experimental text silently at their own pace while their eye movements were recorded. Eye tracking provides accurate information on the time course of the processing: for example, whether readers slow down during the initial encounter of the relevant text information (forward first-pass fixation time), whether they tend to reread sentences during the first-pass reading (first-pass rereadings), or whether they initiate look-backs from subsequent text to reread the relevant text information after the first-pass (look-backs). The different measures reflect different comprehension processes: Forward first-pass fixations reflect the initial processing of text information; first-pass rereadings indicate either integrative processing of the text information or a need to immediately reread the sentence due to, for example, comprehension problems; look-backs reflect the reader's need to restore text information in working memory (e.g., Walczyk & Taylor, 1996).

We coded the think-aloud responses into categories employed in previous studies (Coté & Goldman, 1999; Coté et al., 1998): The categories were associations, self-explanations, comprehension monitoring, paraphrasing, and questions (questions included elaborative questions reflecting the reader's attempts to explain or otherwise elaborate on the text information). We assumed that the think-aloud responses reflecting deeper processing such as self-explanations and questions would be coupled with longer fixation times. Shallower processing, such as associations and paraphrasing, should be preceded by shorter fixation times. Moreover, we tested the hypothesis that the longer processing time observed for perspective relevant as compared to perspective-irrelevant text information (Kaakinen et al., 2002, 2003) is indicative of deeper processing reflecting an increased number of self-explanations and questions.

Participants read an expository text describing four rare diseases of which they had no or very little prior knowledge (Kaakinen et al., 2003). We chose to examine perspective effects in a low prior knowledge condition because we have previously observed a robust perspective effect on both processing and memory for such a text (Kaakinen et al., 2002, 2003). Moreover, think-alouds are more reliable indexes of processing when the text being read is not so easy that comprehension processes are automatized and consequently not available for conscious reporting (e.g., Coté & Goldman, 1999; Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). We have observed that when reading a text about highly familiar text contents, encoding of perspective-relevant information to memory can be a fast and efficient process that does not necessarily require extra processing time (Kaakinen et al., 2003). Thus, perspective effects might not even be detectable with think-alouds for high prior knowledge texts.

METHOD

Participants

Forty-three University of Turku students enrolled in an introductory psychology course participated in the experiment. Participants who had complete data (i.e., fixation time data, think-aloud response data, and recall data) for less than 60% of the relevant or irrelevant target sentences were dropped out. The final data set consisted of measurements from 36 participants (28 women). The age range was 19 to 43 years ($M = 23$ years 9 months, $SD = 4$ years 8 months). All participants received a course credit for participation.

Apparatus

Eye movements were collected by the EYELINK eyetracker manufactured by SR Research Ltd. (Toronto, Ontario, Canada). The eyetracker is an infrared video-based tracking system combined with hyperacuity image processing. There are two cameras mounted on a headband (one for each eye), including two infrared LEDs for illuminating each eye. The headband weighs a total of 450 g. The cameras sample pupil location and pupil size at the rate of 250 Hz. Registration can be done either monocularly or binocularly. We performed it for the selected eye (usually the right eye) by placing the camera and the two infrared lights 4 to 6 cm away from the eye. The resolution of eye position is 15 s of arc, and the spatial accuracy is approximately 0.5°. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the center of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, which are viewed by the head-tracking camera, once the participant sits directly facing the screen. Possible head motion is detected as movements of the four LEDs and is compensated for online from the eye position records.

Materials

The experimental text was adopted from the study of Kaakinen et al. (2003). It was an expository text, 960 words in length, discussing four rare diseases: trigeminusneuralgy, typhus, cystic fibrosis, and scleroderma. Participants had very little knowledge of these diseases before reading the text.[1] The text contained five subtopics, each marked with a subheading ("How to Recognize the Disease," "Origin and Causes," "Treatment," "Prevention," and "Support

[1] After recalling the text, participants were asked to rate how familiar they were with the diseases described in the text on a scale from 1 (*I had never heard about the disease*) to 5 (*I was very familiar with the disease*). The medians for ratings given to trigeminusneuralgy, typhus, scleroderma, and cystic fibrosis were 1, 2, 1, and 2, respectively.

Groups"), and each disease was presented under each subtopic by comparing and contrasting the diseases to each other. The text was written in Finnish, the mother tongue of the study participants.[2]

There were two possible reading perspectives, one of which was assigned to each participant: One perspective was related to trigeminusneuralgy, and the other was related to typhus. Before presenting the text, half of the participants were given the trigeminusneuralgy perspective and the other half the typhus perspective. The reading perspective was induced by instructing the participant to

> Imagine that a close friend of yours has been diagnosed with trigeminusneuralgy/typhus. Everybody is very worried about this common friend, and you have agreed to find out some facts about the disease and to inform the others about it. Read the text in order to be able to do the job.

The text included 10 preselected target sentences relevant to trigeminusneuralgy and 10 relevant to typhus. Because the reading perspective was counterbalanced across participants, for half of the participants the trigeminusneuralgy-relevant target sentences were the perspective-relevant sentences and the typhus-relevant target sentences were the perspective-irrelevant sentences; for the other half, the typhus-relevant sentences were the perspective-relevant sentences and the trigeminusneuralgy-relevant sentences were the perspective-irrelevant sentences. We also matched the target sentences (trigeminusneuralgy vs. typhus) for the mean length and the mean frequencies of the words in the sentences (frequencies were calculated from an unpublished newspaper corpus with the help of the WordMill database program of Laine & Virtanen, 1999).

Procedure

Before the experiment started, the eyetracker was adjusted and calibrated using a nine-point calibration grid that extended over the entire computer screen. The participants were told that they were going to read an expository text, presented on a computer screen, at their own pace and that after some sentences a red asterisk would appear on the screen, which means that they should report whatever they are thinking at the moment. After the initial instructions, thinking-aloud was practiced with an expository text describing a small and remote island. The practice text was 568 words long and was presented on the computer screen sentence by sentence, following the same procedure as used in presenting the experimental text.

Written instructions, in which the reading perspective was given, were presented on the screen before the experimental text. The experimental text was

[2]An English translation of the text is available on request from the authors.

presented sentence by sentence on a computer screen, and the preceding sentences remained on the screen when new sentences appeared. In all, there were eight lines of text on one screen. The participants controlled the presentation of the sentences by pressing a button on a gamepad. A one-point calibration was performed after each screen to correct for the possible changes in the eyetracker settings due to head movements made during thinking aloud. Think-alouds were prompted with a red asterisk after the 10 perspective-relevant and 10 perspective-irrelevant target sentences and after 10 other sentences. Participants did not know in advance exactly when they were supposed to think-aloud—the think-alouds were prompted after the participant had indicated with a button press that he or she was ready to proceed in the text, and the next sentence appeared only when the reader indicated with a button press that he or she had finished with thinking aloud and was ready to continue reading the text. Each button press sends a message to the eyetracker, and the message is visible in the output data. Because the eye movements could not be reliably tracked during thinking aloud due to excessive head movements, eyetracking data recorded during the thinking-aloud phases (i.e., data between the button presses indicating the start and the end of a think-aloud) were removed before the analyses. The session was videotaped and the participants' verbal responses were recorded on a tape recorder. The experimenter was present in the room during the entire session to manage the eye-tracking equipment, video camera, and tape recorder. The experimenter did not prompt the readers to think aloud even if they did not report anything after the prompt (no response was coded as "other"; see the description of the coding of the think-alouds later).

After reading the text, another experiment served as a filler task (about 20 min) before a free recall of the text was collected. Participants were given both verbal and written recall instructions, which stressed that the participants were expected to recall everything they could remember of the text, not just text information related to their reading perspective. At the end of the session, participants were asked to estimate how familiar they were with the diseases described in the experimental text. The total length of the session was approximately 1 hr.

Scoring of the Think-Aloud Protocols

Think-aloud protocols were transcribed from the videotapes (tape recordings were used as a backup) and longer utterances were divided into shorter responses, following the procedure described by Trabasso and Magliano (1996). Only responses produced after the predefined relevant and irrelevant target sentences were further analyzed. Full data (i.e., eye-fixation measures, think-aloud responses, and recall) were available for 642 target sentences (317 relevant and 325 irrelevant), and when the utterances were divided into shorter responses, the total number of responses was 1,052 (the number of responses produced after a target sentence varied between 1 and 8). In selecting the coding categories, we made use of the previous think-aloud studies of Coté and Goldman (1999) and Coté et al. (1998). The following response categories were employed.

1. *Associations* are lateral inferences, which in some sense elaborate the text information but do not improve its coherence; in other words, they are irrelevant to the textbase as well as the situation model ("This reminds me about the famous writer who has multiple sclerosis").

2. *Self*-e*xplanations* include backward inferences that attempt to explain the current text information by using either knowledge retrieved from the reader's prior knowledge or previously given text information ("And this must be because the disease is related to the functioning of the neurons") and summaries that capture the essential information presented in the text but with the reader's own words ("In other words, you need to be very careful not to get a seizure"). This category also included predictions, which are forward inferences that include speculations about the implications of the given text information ("So it would be nearly impossible to fix this problem with an operation").

3. *Questions* indicate that the reader attempts to elaborate on the text information ("How is this kind of surgery performed?") or is otherwise musing on the given information ("Is this the same as … ?").

4. *Monitoring comments* are responses reflecting the reader's own comprehension process or strategies ("Now I don't understand what this means" or "I'm just going quickly through this because this is not about typhus"), as well as comments that imply the reader's emotional involvement with the text ("That's disgusting") or statements that describe the reader's opinion ("That's interesting").

5. *Paraphrases* are simple repetitions of parts of the recently presented text.

6. All other responses were scored as *other*: This category includes responses such as "I'm not thinking about anything," nonresponses, and all other comments that do not fall into the categories described previously.

Two independent raters first divided the utterances into shorter responses, scored 200 responses, and then discussed the exact content of the categories to establish maximally unequivocal definitions of the categories. After this, both raters read and scored the rest of the protocols. Each response was coded into only one category. Ratings were performed independently, and the interrater agreement was 80%. Inconsistencies were resolved through discussion.

RESULTS

We start the analyses by examining the perspective effects. We first report the data for the free recall to see if we can replicate the earlier findings that relevant information is better recalled than irrelevant information. We then turn to the analysis of the different online comprehension measures: fixation times and the think-aloud protocols. Finally, we examine on a more general level how the processing time measures are related to the think-aloud responses by collapsing the data across relevant and irrelevant target sentences.

Perspective Effects on Recall

The recall protocols were scored for the gist of the meaning of the target sentences. Two independent raters scored the protocols, and the interrater agreement was 95%. A third rater resolved the inconsistencies. The data were analyzed with a repeated-measures analysis of variance (ANOVA) with relevance as the within-participant factor. As can be seen in Table 1, readers recalled more perspective-relevant than perspective-irrelevant text information, $F(1, 35) = 25.30$, $MSE = .03$, $p < .001$, $\eta^2 = .42$. Thus, we replicated the previous findings that a reading perspective guides readers to construct a perspective-bound representation of the text. The critical question is whether the perspective effect is also seen in the processing measures: fixation times and think-aloud responses.

Perspective Effects on Fixation Time Measures

Three processing time measures were derived from readers' eye fixation patterns: (a) first-pass progressive fixation time, (b) first-pass rereading time, and (c) look-back time (see Hyönä et al., 2003). All fixations landing on a target segment during the initial reading of the sentence are defined as first-pass fixations, whereas fixations returning back to a target sentence from a subsequent sentence

TABLE 1
The Mean Fixation Time and the Mean Recall Rate for the Relevant and Irrelevant Target Sentences

	Target Sentence			
	Relevant		*Irrelevant*	
Measure	*M*	*SE*	*M*	*SE*
First-pass progressive fixation time[a]	34.39	1.51	32.19	1.48
First-pass rereading time[a]	24.42	2.52	17.05	1.63
Look-back time[a]	4.85	1.31	3.65	1.05
Recall (%)	45.17	3.44	24.52	3.63

Note. $N = 36$.
[a]Milliseconds per character.

are called look-back fixations. First-pass fixations were further divided into progressive first-pass fixations and first-pass rereadings. Progressive fixations are forward-going fixations that land on unread parts of a sentence and are thought to index the most immediate processing. First-pass rereading time is the summed duration of fixations landing on the read parts of a sentence during the first-pass reading and reflects the reader's immediate need to reread a sentence. First-pass rereadings reflect attempts to "buy time" to build strong enough memory links for text information (i.e., integrate it to the developing memory representation) before proceeding to the next sentence. Finally, the look-back time is a measure of later processing: Look-backs are fixations returning to a sentence from subsequent sentences, and they serve the purpose of restoring text information in working memory. The mean fixation times per character were computed for each sentence by dividing the first-pass progressive fixation time, first-pass rereading time, and look-back time by the number of characters in the sentence. These mean per character fixation times were averaged separately for the 10 perspective-relevant and 10 perspective-irrelevant target sentences. The data were analyzed with repeated-measures ANOVAs with relevance as the within-participant factor.

As can be seen in Table 1, a perspective effect (i.e., longer processing time for relevant than for irrelevant information) was observed in the *progressive first-pass fixation time*, $F(1, 35) = 6.23$, $MSE = 13.92$, $p = .017$, $\eta^2 = .15$. In other words, readers invested more attentional resources to relevant text information and slowed down the pace of reading immediately when they encountered relevant information in the text.

Relevant sentences were also *reread during the first-pass* more than perspective-irrelevant sentences, $F(1, 35) = 13.31$, $MSE = 73.56$, $p = .001$, $\eta^2 = .28$. Thus, in addition to slowing down the initial processing of text information during the first-pass reading of the sentences, readers also reread the relevant sentences more than irrelevant sentences.

The small difference between the *look-back times* to relevant and irrelevant sentences failed to reach significance, $F(1, 35) = 2.12$, $MSE = 12.35$, $p = .154$, $\eta^2 = .06$. This result is different from what we have observed before (Kaakinen et al., 2002, 2003), and it indicates that the think-aloud procedure may have influenced the way readers inspect the text by reducing the need to look back in text.

Perspective Effects on Think-Aloud Responses

The fixation time measures, more specifically the first-pass progressive fixation and rereading times, revealed a clear effect of a reading perspective on how readers process relevant and irrelevant information in text. Thus, we expected to observe a perspective effect in the other online processing measure employed in the study: think-aloud responses.

TABLE 2
The Raw Frequencies, Number of the Participants, and the Probabilities of the Different Think-Aloud Responses to Relevant and Irrelevant Target Sentences

	Target Sentence							
	Relevant				*Irrelevant*			
Response Category	*f*	*n*	*P*	*SE*	*f*	*n*	*P*	*SE*
Association	74	27	.23	.04	80	25	.25	.05
Self-explanation	72	25	.22	.04	77	24	.24	.05
Paraphrase	125	28	.38*	.07	80	30	.24	.03
Monitoring	157	34	.49	.05	164	35	.50	.05
Question	74	26	.23	.05	62	20	.20	.05
Other	43	17	.15	.04	44	15	.14	.04

Note. f = raw frequency; n = number of participants making the type of response; P = probability of making a response; SE = standard error of the probability. Data was available from 317 relevant and 325 irrelevant target sentences.
$^*p < .05$.

We computed the probability that a participant makes a specific type of think-aloud response after relevant and irrelevant target sentences by dividing the number of the responses by the number of the available target sentences.[3] For example, if there were data from eight relevant and nine irrelevant target sentences in the data set for a participant, and the participant made seven paraphrases after relevant and one paraphrase after irrelevant target sentences, the probability of a paraphrase would be 7/8 = .88 and 1/9 = .11 for the relevant and irrelevant target sentences, respectively. The probabilities are presented in Table 2.

First, we computed a 2 (relevance) × 5 (think-aloud response type) repeated-measures ANOVA to examine if there were differences between the probabilities of responses to the relevant and irrelevant target sentences. The "other" category was left out from the analyses because it is theoretically uninteresting. Greenhouse–Geisser corrections for the degrees of freedom were used when appropriate. The think-aloud responses were not evenly distributed across the different response categories, as indicated by a significant main effect of think-aloud response type, $F(4, 140) = 8.06$, $p < .001$, $\eta^2 = .19$. The critical two-way interaction between relevance and think-aloud response type failed to

[3]The think-aloud data was available from 642 target sentences: 317 relevant and 325 irrelevant. Data was missing from 35 irrelevant and 43 relevant target sentences; 16 participants had some missing data for irrelevant and 24 participants for the relevant targets. Thus, the raw frequencies had to be corrected for the missing data before the data could be analyzed.

reach significance, $F(3, 107) = 1.80$, $p = .15$, $\eta^2 = .05$. In other words, relevant and irrelevant text information probed overall very similar responses. Separate Wilcoxon signed ranks tests for each response category showed that only paraphrases were more frequent after relevant than after irrelevant target sentences, $p = .027$.

Combining Think-Aloud Responses With Fixation Time Measures

Finally, we were interested in the correlation between the different fixation time measures and the think-aloud response types. As all participants did not produce all types of think-aloud responses, participant means could not be computed as in the previous analyses. Instead, we collapsed the data across relevant and irrelevant target sentences and treated the 1,052 responses as a single data set. We first tested for overall differences in fixation times between the response categories. Significant overall differences were further examined by comparisons between two groups of responses: responses reflecting deeper processing (combination of self-explanations and questions) and responses reflecting shallower processing (combination of associations and paraphrases). The "other" responses were left out from all analyses. Because the numbers of the utterances in the response categories were grossly unequal, we chose to use nonparametric statistics.

First, the Kruskall–Wallis test was used to test for the differences in the fixation time measures between the different response types. The results showed that there were differences in the progressive first-pass fixation time between the response types, $\chi^2(4, N = 965) = 19.96$, $p = .001$. As can be seen in Panel A of Figure 1, questions were preceded by longer progressive fixation times than any of the other responses. On the other hand, associations were coupled with relatively short progressive fixation times. Next, we examined with a Mann–Whitney U test whether deeper processing, as indexed by self-explanations and questions ($M = 36.30$, $SE = .96$), were coupled with longer progressive first-pass fixation times than shallower processing as indicated by paraphrasing and associations ($M = 32.69$, $SE = .64$). Indeed, such a difference was observed in the progressive first-pass fixation times preceding these two types of responses, $p = .014$.

The analysis of the first-pass rereading time also revealed overall differences between the response types, $\chi^2(4, N = 965) = 31.58$, $p < .001$. It is evident from Panel B of Figure 1 that questions were again coupled with longer fixation times than any other response type. Also, self-explanations were preceded by relatively long first-pass rereading times, whereas associations, paraphrases, and monitoring comments were preceded by approximately equally short first-pass rereading

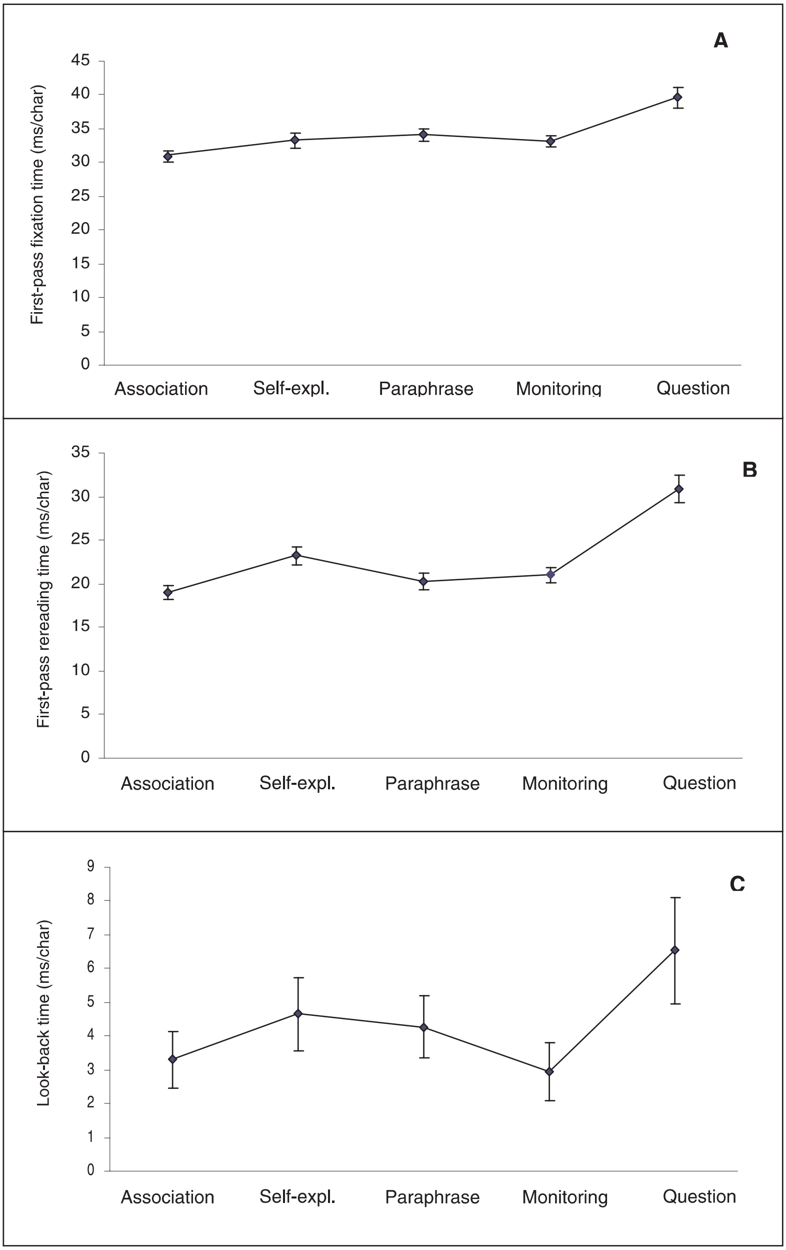

FIGURE 1 The mean first-pass progressive fixation time (Panel A), first-pass rereading time (Panel B), and look-back time (Panel C) for the different response types. Error bars represent standard errors of the means. Self-expl. = self-explanation.

times. Deeper processing was coupled with longer first-pass rereading time (M = 26.86, SE = 1.50) than shallower processing (M = 19.71, SE = 1.28), $p < .001$.[4]

The number of look-backs was very low in general, and thus it was not surprising that there was no main effect of response type, $\chi^2(4, N = 965) = 5.41$, $p = .248$. Panel C of Figure 1 presents the mean look-back times for the different response types.

DISCUSSION

The purpose of this study was threefold. First, we aimed to replicate our previous findings showing that the perspective-relevant sentences attract longer fixation times than perspective-irrelevant sentences. We were particularly interested in whether the perspective effect is observed in the initial processing of the sentences. Second, we were interested in whether the longer processing times observed for relevant sentences reflect deeper processing invested in these sentences as measured by the think-aloud responses. Third, we examined the validity of the think-aloud protocols by simultaneously combining eyetracking and think-aloud protocols. We next discuss each of these issues in more detail.

Our results replicated the previous findings (Kaakinen et al., 2001, 2002, 2003) that the perspective-relevant information is both processed longer and also recalled better than perspective-irrelevant information. In particular, the perspective effect in processing was detected during the initial reading of the sentences, as indexed by the progressive first-pass fixations. In addition, a reading perspective increased the immediate rereading of the relevant text information. These findings show that, in addition to a lower level control of eye movements in reading, there is also a strong top-down component involved (see also Wiley & Rayner, 2000). The current models of eye-movement control in reading emphasize the word-level processes (e.g., Reichle, Rayner, & Pollatsek, 2003), and the higher level guidance has so far been left aside. Clearly, in the future attention needs to be paid to the higher level processes as well.

In our previous studies (Kaakinen et al., 2002, 2003), we have also observed more look-backs to perspective-relevant than to perspective-irrelevant sentences.

---

[4]The processing time differences observed between the deep and shallow response types were not caused by a general processing time difference between the relevant and irrelevant sentences. First of all, there were no differences in the amount of reported deep versus shallow processing on relevant and irrelevant text information both ts < 2. Second, when we computed 2 (relevance) × 2 (depth of processing) repeated-measures ANOVAs for the different processing time measures, none of the Relevance × Depth of Processing interactions proved significant, largest $F(1, 652) = 1.02$, $p = .313$, $\eta^2 = .002$ for the first-pass rereadings. Only the main effects of relevance and depth of processing were significant for both first-pass reading measures, smallest $F(1, 652) = 5.03$, $p = .025$, $\eta^2 = .008$ for the main effect of relevance on the progressive first-pass fixation times.

In this study we failed to replicate the perspective effect in the look-back times. In fact, participants rarely returned to read sentences. A study by Coté et al. (1998) suggested that the reduction in the number of look-backs may be a byproduct of the think-aloud procedure. Coté et al. compared think-aloud reading to silent reading and found that readers were engaged in less overt looking back when thinking aloud. They concluded that the think-aloud task itself may increase the attentional effort readers invest on the text so that look-backs are not as necessary as in the silent reading. Our finding corroborates this interpretation.

Do the longer first-pass reading times on the perspective-relevant sentences in comparison to the perspective-irrelevant sentences reflect deeper processing devoted to relevant text information? We failed to find support for this hypothesis. Neither self-explaining nor elaborative questions were more probable after relevant than after irrelevant text. Instead, readers tended to paraphrase more parts of relevant text. Repeating the text information serves the purpose of holding the relevant information in the working memory for a longer time (Pressley & Afflerbach, 1995). Keeping the text information active in working memory helps to connect it to the developing textbase. Our results thus suggest that a reading perspective did not encourage readers to construct deeper understanding of the relevant text information during the course of reading. Instead, readers were satisfied with constructing a good textbase for the relevant text. This finding is consistent with previous memory studies showing that readers usually are better in recalling relevant than irrelevant facts of the text (e.g., Baillet & Keenan, 1986). The free-recall task employed in these studies is thought to mainly tap into the textbase or surface level understanding of the text (McNamara, Kintsch, Songer, & Kintsch, 1996).

Previous research shows that readers often use explanatory inferences to construct a coherent and comprehensive memory representation of a text (e.g., Coté et al., 1998; Coté & Goldman, 1999; Magliano, Trabasso, & Graesser, 1999; Trabasso & Magliano, 1996; Trabasso & Suh, 1993; van den Broek et al., 2001). However, in this study, readers were not extensively involved in these kinds of activities when they encountered relevant information in the text, even though they had better memory for relevant than for irrelevant text information. We can think of three different factors that may have contributed to this finding: readers' lack of sufficient prior knowledge, type of text, and task demands.

First, it is quite plausible that explaining was simply too difficult because readers had no or minimal prior knowledge that could be brought to bear in aiding comprehension (Chiesi, Spilich, & Voss, 1979; McNamara & Kintsch, 1996). There is evidence showing that explaining is possible even in low prior knowledge conditions but only if special instructions, emphasizing explaining, are given to the readers (McNamara, 2004). However, we think that it is not likely that the readers would spontaneously produce explanations if they lack the sufficient knowledge, especially when reading a relatively difficult expository text with the instructions to "be able to tell other people about the disease." Second, the type of text used in this study may have diminished the number of explanations. Our relatively descriptive expository text may not have invited as many explanations as narratives or expository texts that have a clear causal chain structure. Third, the task the readers were performing (to be able to tell the others facts about an imaginary friend's disease) may not have invited rigorous situation model construction involving elaborative inferencing. Instead, the task combined with the fact that the readers had minimal knowledge of the text contents and with the type of text structure may have primarily emphasized textbase construction.

Our results support the validity of the think-aloud protocols as a measure of online comprehension. Responses that are thought to reflect deeper level comprehension activities (self-explaining and elaborative questioning) were coupled with longer eye fixation times than responses indexing shallower processing (associations and paraphrasing). In other words, the type of processing indicated by the reader's verbal response was related to the amount of processing time invested in the probed sentence. This correlation was observed for the progressive first-pass fixation times, suggesting that the "deeper" processing was truly done online, that is, during the initial encounter with a sentence. To our knowledge, this study is the first to have simultaneously collected eye movement and think-aloud data, thus providing unique information about the validity of the think-aloud protocols as measures of text comprehension. However, our results are restricted to one type of expository text and specific reading instructions. In the future, it will be important to examine whether the correlation between the verbal reports and online processing times holds across different texts and reading situations.

ACKNOWLEDGMENTS

Johanna K. Kaakinen is now at the Florida State University. Jukka Hyönä acknowledges the financial support of Suomen Akatemia (the Academy of Finland). We thank Annu Haapakangas and Anna-Mari Nurminen for their help in the data analyses. We are also grateful to Susan Goldman, Joe Magliano, Nathalie Coté, and an anonymous reviewer for their useful and constructive criticism.

REFERENCES

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17,* 1–12.

Anderson, R. C., Pichert, J. W., & Shirey, L. L. (1983). Effects of the reader's schema at different points in time. *Journal of Educational Psychology, 75,* 271–279.

Baillet, S. D., & Keenan, J. M. (1986). The role of encoding and retrieval processes in the recall of text. *Discourse Processes, 9,* 247–268.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18,* 257–274.

Coté, N., & Goldman, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 169–193). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25,* 1–53.

Ericsson, K. A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fletcher, C. R. (1986). Strategies for the allocation of short-term memory during comprehension. *Journal of Memory and Language, 25,* 43–58.

Goetz, E. T., Schallert, D. L., Reynolds, R. E., & Radin, D. I. (1983). Reading in perspective: What real cops and pretend burglars look for in a story. *Journal of Educational Psychology, 75,* 500–510.

Goldman, S. R., Varma, S., & Coté, N. (1996). Extending capacity-constrained construction-integration: Toward "smarter" and flexible models of text comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 73–113). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Hyönä, J., Lorch, R. F., Jr., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2001). Individual differences in perspective effects on text memory. *Current Psychology Letters: Behaviour, Brain & Cognition, 5,* 21–32.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective effects on on-line text processing. *Discourse Processes, 33,* 159–173.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 447–457.

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

Laine, M., & Virtanen, P. (1999). WordMill lexical search program [Computer software]. Center for Cognitive Neuroscience, University of Turku, Finland.

Lehman, S., & Schraw, G. (2002). Effects of coherence and relevance on shallow and deep text processing. *Journal of Educational Psychology, 94,* 738–750.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91,* 615–629.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38,* 1–30.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22,* 247–288.

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,* 1–43.

Moravcsik, J. E., & Kintsch, W. (1995). Writing quality, reading skills, and domain knowledge as factors in text comprehension. In J. M. Henderson, M. Singer, & F. Ferreira (Eds.), *Reading and language processing* (pp. 232–246). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Olson, G. M., Duffy, S. A., & Mack, R. L. (1984). Thinking out-loud as a method of studying real-time comprehension processes. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 253–286). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pichert, J. W., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology, 69,* 309–315.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–526.

Rothkopf, E. Z., & Billington, M. J. (1979). Goal-guided learning from text: Inferring a descriptive processing model from inspection times and eye movements. *Journal of Educational Psychology, 71,* 310–327.

Trabasso, T., & Magliano, J. (1996). Conscious understanding during text comprehension. *Discourse Processes, 21,* 255–288.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes, 16,* 3–34.

van den Broek, P., Lorch, R. F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29,* 1081–1087.

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

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* San Diego, CA: Academic.

Walczyk, J. J., & Taylor, R. W. (1996). How do the efficiencies of reading subcomponents relate to looking back in text? *Journal of Educational Psychology, 88,* 537–545.

Wiley, J., & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words: Evidence from eye movements. *Memory & Cognition, 28,* 1011–1021.