

# Getting a Clue: How Initially Attending to a Picture Fosters Learning from Subsequent Text

Alexander Eitel, Katharina Scheiter, Anne Schüller  
Knowledge Media Research Center  
Konrad-Adenauer Str. 40, Tuebingen, Germany  
{a.eitel, k.scheiter, a.schueler}@iwm-kmrc.de

**Abstract.** When learning from the simultaneous presentation of instructional text and pictures, the picture is sometimes shortly attended before the text is read. A prior study revealed that shortly attending to a picture of a schematic causal system leads to the extraction of both its gist (< 150 ms) and its functioning (< 600 ms). In the current study it is investigated whether extracting these kinds of information from a picture of a causal system prior to reading the instructional text fosters learning. It is hypothesized that getting the gist and the functioning of a causal system enhances comprehension of subsequent text. Further, we assume that it is more effective to present the picture of a causal system for the time it takes to extract its functioning (600 ms) than to extract its gist (150 ms) prior to reading. These assumptions are currently tested in a 5x1 between-subjects design.

**Keywords:** Gist; functioning; learning from text and pictures; short presentation times

## Theory

Research in multimedia has repeatedly shown that learning from text and pictures results in higher comprehension than learning from text alone (Mayer, 1989). However, there has been little research about how pictures are processed during learning from text and pictures with a few noteworthy exceptions. In an early study by Hegarty and Just (1993), learners first processed text information, and then switched to the picture probably in order to integrate the information from both sources. On the other hand, a study from Stone and Glock (1981) showed that learners initially processed the picture for a short time (1000 – 2000 ms) before they started to read the text. According to the authors, subjects initially looked at the picture in order to get a first impression (i.e., gist) of what the material was about. However, it is yet unclear what role looking briefly at a picture prior to reading a text may play for understanding the presented content.

Prior to identifying the role of briefly attending to the picture before reading the respective text, it has to be investigated what is actually extracted from briefly viewing a picture without text. In scene perception research, it is a well known finding that the gist (i.e., general theme) of a picture of a scene is rapidly extracted (e.g., 150 ms; Oliva & Schyns, 2000). The recognition of details, however, is poor after briefly looking at the picture of a scene (250ms; Liu & Jiang, 2005), but it improves with slightly longer presentation times (1000 ms; Loftus, Nelson & Kallman, 1983). Hence, the gist of a scene is rapidly extracted without the need to identify single details.

## Prior Study

In a prior study, we investigated whether findings from scene perception research can be applied to instructional material. Since causal systems were frequently used as instructional material in studies on learning from text and picture (e.g., Mayer, 1989), the extraction of information from briefly attending to causal systems was investigated. One hundred sixty schematic and photorealistic pictures of scenes and causal systems were presented at four different presentation times (150, 600, 2000, 6000 ms) to subjects. Scenes depicted everyday situations. A causal system always had a certain purpose (e.g., pulling weight). It consisted of multiple components, where at least one component was influenced by another component – hence, removing one component would have changed the functioning of the system. Therefore, in causal systems, not only the ability to extract information about gist and details after brief exposure was assessed, but also understanding of the functioning of the system. The results

demonstrate that the gist was rapidly extracted (< 150 ms) in both scenes and causal systems, confirming prior research from gist extraction in scenes (e.g., Oliva & Schyns, 2000) and expanding it to instructional material. Moreover, details were recognized more accurately at longer presentation times, which is in line with prior research from detail extraction in scenes (Loftus et al., 1983). Comprehension of the functioning quickly reached an asymptote in schematic pictures of causal systems (at 600 ms). Hence, subjects' initial glance at the schematic picture of instructional material for 1000 to 2000 ms in the study of Stone and Glock (1981) presumably not only served the purpose of extracting the gist, but also of getting a preliminary understanding of the functioning of the system.

### *Research Question of Current Study*

Learners initially attend to the picture for a short time prior to reading when learning from text and pictures (Stone & Glock, 1981). Even though the role of the picture in this process is still unclear, the prior study revealed that within this brief attention devoted to the picture (1000 to 2000 ms), information both about the gist (< 150 ms) and about the functioning (< 600 ms) of the instructional picture (causal system) can be extracted. In the current study, it is investigated whether initially attending to a picture of a causal system is effective for learning, and whether extraction of its gist has a different effect on learning from subsequent text than extraction of its functioning.

We expect higher learning outcomes when the picture of a causal system is presented for the time to extract its gist (150 ms) prior to the text ('gist before text') than presenting text alone. This is hypothesized, because the gist of a picture provides a mental scaffold (Friedman, 1979) that can be held in memory for some time (Castelhano & Henderson, 2007). Hence, subsequent text information possibly is added to that scaffold, which leads to incremental mental model construction (Hegarty & Just, 1993). An integrated mental model of text and picture information (Schnotz, 2002) should be the result that, in turn, should be measurable in terms of higher comprehension and recall of the learning content compared to learning from text alone. Moreover, text passages about the spatial arrangement of the components within the picture should be read more fluently in the 'gist-before-text' than in the 'text-alone' condition, since learners already have a mental picture in the former condition.

Furthermore, we expect higher comprehension of the text when the picture (of a causal system) is presented for the time it takes to extract its functioning (600 ms; 'functioning before text') than when presented for the time it takes to extract its gist (150 ms). This is due to the assumption that the extraction of the functioning possibly leads to a functional scaffold, which means that a preliminary understanding of the functioning is acquired and subsequent text information regarding the functioning can be added to this scaffold. This should lead to an overall higher comprehension of the learning content and to a more fluent reading of text passages where the functioning of the causal system is described in the 'functioning before text' than in the 'gist before text' condition.

To measure the overall effectiveness of a short presentation of the picture of a causal system before a respective instructional text, the two conditions ('gist before text', 'functioning before text') are not only compared to a 'text alone' condition but also to the self-paced presentation of the picture of a causal system before the text ('self-paced before text'). Since learners attend to pictures only for a short time before reading text (Stone & Glock, 1981), and assuming that this is an effective strategy we expect learning outcomes to be as high in the 'functioning before text' as in the 'self-paced before text' condition. Thus, we expect higher learning outcomes in all three conditions in which a picture is presented before the text than in the 'text alone' condition. Finally, all four conditions are compared to the simultaneous presentation of the text and the picture, since this type of presentation has been used

most frequently in studies on learning with text and picture (e.g., Mayer, 1989). The simultaneous presentation of text and picture is expected to lead to the highest learning outcomes, since the integration of text and picture information that takes place after reading, and that has shown to foster learning (Hegarty & Just, 1993), can only be realized in this condition.

## Method

The current study followed a 5x1 between-subjects design (see Figure 1). Instructional material from three different domains (i.e., human heart, hydraulic drum brakes, and pulley systems) was presented to subjects. The pictures of these three domains had been used as schematic causal systems in the prior study, and both their gist had been extracted to a high extent at 150 ms and their functioning had been understood quite well at 600 ms. Learning outcomes in the current study were assessed in terms of retention, free recall, and transfer/comprehension measures after the learning phase in each of the five experimental conditions. Furthermore, eye tracking methodology was used to assess how the subsequently read text was processed, and how processing of the text differed between conditions.

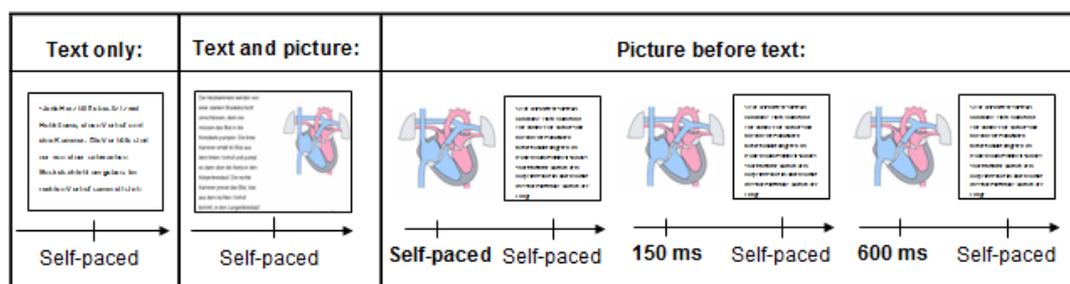


Figure 1: Overview of the five experimental conditions.

Data collecting is currently on the way. Results will be presented at the conference.

## References

- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human*, 33, 753-763.
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 108, 316-355.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717-742.
- Liu, K., & Jiang, Y. (2005). Visual working memory for briefly presented scenes. *Journal of Vision*, 5, 650-658.
- Loftus, G. R., Nelson, W. W., & Kallman, H. J. (1983). Differential acquisition rates for different types of information from pictures. *Quarterly Journal of Experimental Psychology-A*, 35, 187-198.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81, 240-246.
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41, 176-210.
- Schnotz, W. (2002). Wissenserwerb mit Multimedia. *Unterrichtswissenschaft*, 4, 293-318.
- Stone, D. E., & Glock, M. E. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology*, 73, 419-426.