Pergamon

Learning and Instruction 13 (2003) 141–156

Learning and Instruction

www.elsevier.com/locate/learninstruc

Construction and interference in learning from multiple representation

Wolfgang Schnotz *, Maria Bannert

Department of General and Educational Psychology, University of Koblenz-Landau, Thomas-Nast Str. 44, 76829 Landau, Germany

Abstract

This paper presents an integrated view of learning from verbal and pictorial representations. Learning from these representations is considered as a task oriented process of constructing multiple mental representations. Construction of these representations includes information selection and information organisation, parsing of symbol structures, mapping of analog structures as well as model construction and model inspection. Based on this theoretical view an experiment was conducted to analyse the effects of different kinds of multiple external representations on the structure of mental models. Sixty university students were randomly assigned to one of the three experimental conditions. The text-only group learned the subject matter with a hypertext, whereas the other two groups learned the subject matter with a hypermedium including this hypertext and different kinds of graphics. The findings indicate that the structure of graphics affects the structure of the mental model. They also indicate that presenting graphics is not always beneficial for the acquisition of knowledge. Whereas task-appropriate graphics may support learning, task-inappropriate graphics may interfere with mental model construction.

Keywords: Descriptive representations; Depictive representations; Text comprehension; Picture comprehension; Propositional representations; Visual images; Mental models; Support of mental model construction; Interference with mental model construction

* Corresponding author. Tel.: +49-6341-990-234; fax: +49-6341-990-240.
E-mail address: schnotz@uni-landau.de (W. Schnotz).

0959-4752/03/$ - see front matter © 2003 Published by Elsevier Science Ltd.
doi:10.1016/S0959-4752(02)00017-8

1. Introduction

Multimedia allow flexible combinations of different forms of external representations like written or spoken texts, static or animated pictures, sound and music (van Someren, Reimann, Boshuizen, & de Jong, 1998). Research on learning from multiple external representations has focused primarily on combinations of texts and pictures. Its main finding was that text information is remembered better when it is illustrated by pictures than when there is no illustration (Levie & Lentz, 1982; Levin, Anglin, & Carney, 1987). These findings were usually explained with Paivio's dual coding theory (Paivio, 1986Paivio, 1986; Clark & Paivio, 1991Clark and Paivio, 1991). According to this theory, verbal information and pictorial information are processed in different cognitive subsystems: a verbal system and an imagery system. Words and sentences are usually processed and encoded only in the verbal system, whereas pictures are processed and encoded both in the imagery system and in the verbal system. Thus, the memory-enhancing effect of pictures in texts is ascribed to the advantage of a dual coding as compared to single coding in memory.

Mayer (1997) has elaborated the dual coding theory in order to explain why pictures support under specific conditions the understanding of technical or physical phenomena. Mayer assumes that verbal and pictorial explanations are processed in different cognitive subsystems and that they result in the construction of different mental models. Verbal selection processes lead to a propositional text base and verbal organisation processes result in a text-based mental model. Similarly, pictorial selection processes lead to an image base and pictorial organisation processes result in a picture-based mental model. The verbal organisation processes take place in the verbal part of working memory, whereas the pictorial organisation processes occur in the pictorial part of working memory (Baddeley, 1992; Chandler & Sweller, 1991). The text-based model and the picture-based model are then integrated in a one-to-one mapping process. During this mapping process elements of the text-based model are mapped on elements of the picture-based model, and vice versa. Similarly, relations within the text-based model are mapped on relations within the picture-based model, and vice versa. Integration requires that components of the text-based model and corresponding components of the picture-based model are simultaneously activated in working memory.

In a series of experiments Mayer and his co-workers found that pictures support comprehension when texts and pictures are explanatory, when verbal and pictorial content are related to each other, when verbal and pictorial information are presented closely together in space or time and when individuals have low prior knowledge about the subject domain but high spatial cognitive abilities (Mayer, 1997). The parallelism between text processing and picture processing assumed in his model seems to be questionable, however, because texts and pictures use different sign systems resulting in fundamentally different forms of representations which are referred to as descriptive and depictive representations and which have different uses for different purposes: Whereas descriptions are more powerful in representing different kinds of subject matter, depictions are better suited to draw inferences (cf. Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991).

2. Descriptive and depictive representations

Spoken or written texts, mathematical equations and logical expressions, for example, are descriptive representations. A descriptive representation consists of symbols describing an object. Symbols have (like e.g. the words in natural language) an arbitary structure, and they are related to the content they represent by means of a convention (cf. Peirce, 1906). In a sentence like *"The earth rotates around its axis"*, nouns are used to refer to entities, whereas verbs and prepositions are used to relate these entities to each other. Accordingly, descriptive representations contain signs for relations.

Pictures, sculptures, or physical models, for example, are depictive representations. A depictive representation consists of iconic signs. Although depictive representations allow us to extract relational information, they do not contain symbols for these relations. Instead, they possess specific inherent structural features that allow us to read off relational information, and they are associated with the content they represent through these common structural characteristics. Examples of depictive representations are provided in Fig. 1. The two pictures illustrate the fact that there exist on earth different daytimes and dates at the same time. The picture at the top shows the earth as a rectangle moving across a time axis like a kind of flying carpet. In this picture, the left–right dimension is used to represent earlier–later relations in time. The picture at the bottom shows the earth as a circle (or sphere) seen from the North Pole, which rotates in a bowl of different daytimes and dates. In this picture, counter-clockwise rotation is used to represent earlier–later relations in time.

The distinction between descriptions and depictions applies also to internal mental representations. On the one hand, the mental representation of a text surface constructed during the processing of a spoken or written text and the propositional representation of the text contents are internal descriptions, since they describe the represented object with the help of symbols. On the other hand, visual images and mental models are internal depictions, since inherent structural characteristics are used here for the purpose of representation. Visual images are perception-proximal representations, since visual images and visual perceptions are based on the same cognitive mechanisms (Kosslyn, 1994). Mental models, on the contrary, are not sensorically specific. A mental model of a spatial configuration, for example, can be constructed not only by visual perception, but also by auditive or by kinaesthetic or by haptic perception.

Since mental models are not bound to specific sensoric modalities, they are more abstract than perceptual images. On the one hand, a mental model contains less information than the corresponding visual image, because irrelevant details may be omitted. On the other hand, a mental model contains more information than the corresponding visual image, because it includes also prior knowledge information, which is not present in the visual perception. For example, a mental model of a technical device like a brake can contain information about causal relationships, which are not explicitly included in the corresponding picture of the brake (cf. Mayer & Anderson, 1992).

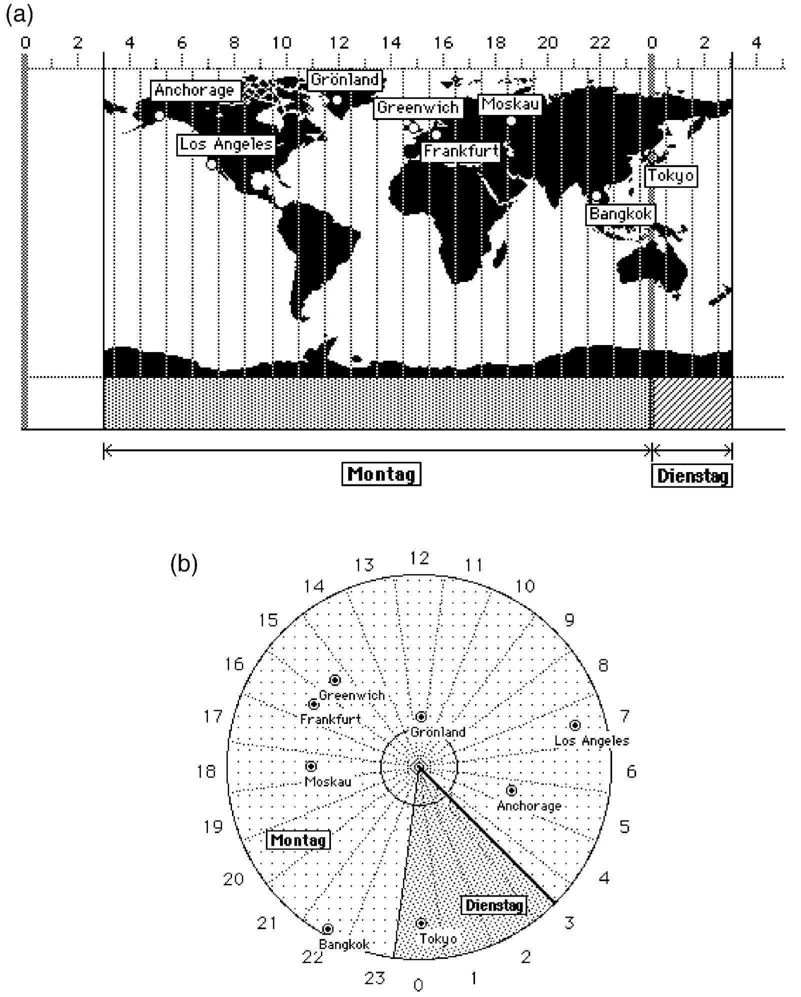

Fig. 1. Visualization of geographic time differences by a carpet diagram (top) and by a circle diagram (bottom).

3. A cognitive model of multimedia learning

Based on the distinction of descriptions and depictions as fundamentally different forms of representations we argue for an alternative model of learning from text and pictures. An outline of this model is shown in Fig. 2. The model consists of a descriptive (left side) and a depictive (right side) branch of representations. The descriptive branch comprises the (external) text, the (internal) mental representation of the text surface structure and the (internal) propositional representation of the semantic content. The interaction between these descriptive representations is based on symbol processing. The depictive branch comprises the (external) picture or diagram, the

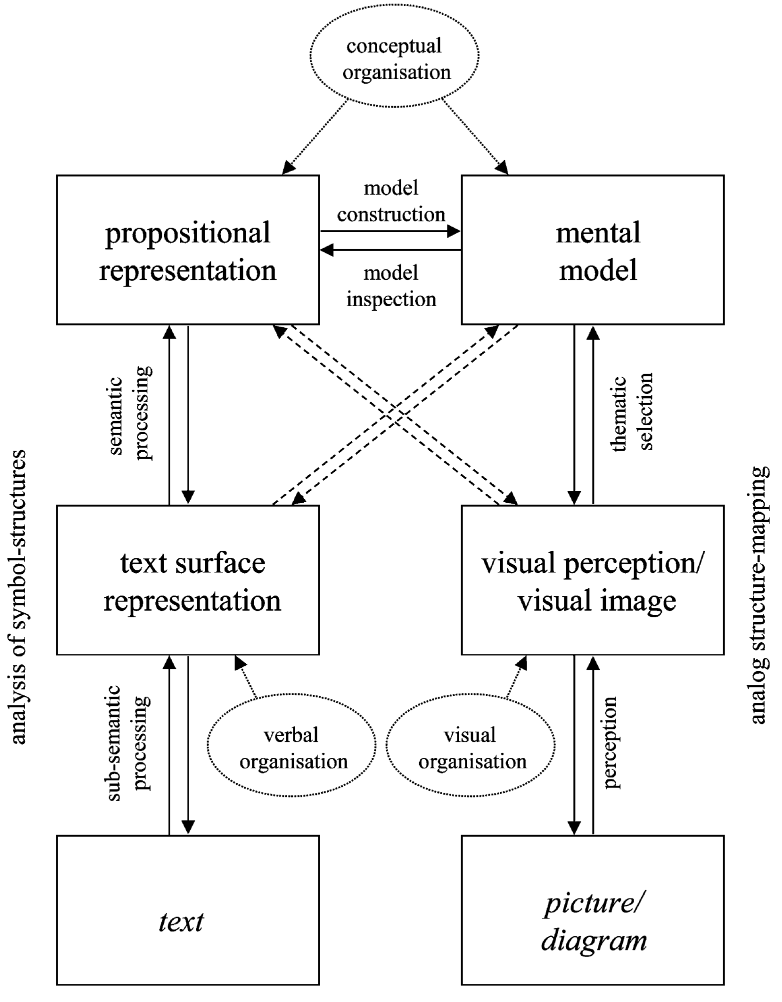

Fig. 2. An integrated model of text and picture comprehension.

(internal) visual perception or image of the graphical display and the (internal) mental model of the depicted subject matter. The interaction between these depictive representations is based on processes of analog structure mapping due to structural correspondences (Gentner, 1989).

According to this model, the reader of a text constructs a mental representation of the text surface structure, generates a propositional representation of the semantic content (i.e. a text base), and finally constructs from the text base a mental model of the subject matter described in the text (van Dijk & Kintsch, 1983; Schnotz, 1994; Weaver, Mannes, & Fletcher, 1995). These construction processes are based on an interaction of bottom-up and top-down activation of cognitive schemata, which have both a selective and an organizing function. Task-relevant information is selected through top-down activation, and the selected information is then organized into a coherent mental representation of the text surface structure. Processes of conceptual organization starting from the text surface representation result in a coherent propositional representation, which in turn triggers the construction of a mental model. Mental model construction implies a transition from a descriptive to a depictive representation. Propositional representations and mental models are assumed to interact continuously via processes of model construction and model inspection guided by cognitive schemata. Based on the propositional information and the default values of the schemata, the mental model is constructed through gestalt-directed composition rules from some depictive primitives in a way to represent a typical instance of what is described in the text. After a mental model has been constructed, schema-directed processes of model inspection can be applied in order to read off new information from the model. This information is encoded in a propositional format and, thus, elaborates the propositional representation which can be externalised by verbal utterances.

In picture comprehension, the individual first creates a visual mental representation of the picture's graphic display through perceptual processing and then constructs a mental model as well as a propositional representation of the subject matter shown in the picture through semantic processing. In perceptual processing, task-relevant information is selected through top-down activation of cognitive schemata and then visually organized through automated visual routines (Ullmann, 1984). Perceptual processing includes identification and discrimination of graphic entities as well as the visual organization of these entities according to the Gestalt laws (Wertheimer, 1938; Winn, 1994). The resulting visual perception created as a surface representation of the picture is an internal depictive representation. It retains structural characteristics of the picture, and it is sensorically specific because it is linked to the visual modality. As perception and imagery are based on the same cognitive mechanisms, the same kind of representation can also be referred to as a visual image, if it is created on the basis of internal world knowledge rather than external sensory data (cf. Kosslyn, 1994; Shepard, 1984).

In order to understand a picture rather than only to perceive it, semantic processing is required. The individual has to construct a mental model of the depicted subject matter through a schema-driven mapping process, in which graphic entities are mapped onto mental entities and in which spatial relations are mapped onto semantic relations as encoded in the mental model. In other words, picture comprehension is considered as a process of analogical structure mapping between a system of visuo-spatial relations and a system of semantic relations (cf. Falkenhainer, Forbus, & Gentner, 1989/90; Schnotz, 1993). This mapping can take place in both directions: It is possible to construct a mental model bottom-up from a picture, and it is also possible to evaluate an existing mental model top-down with a picture. In understanding realistic pictures the individual can use cognitive schemata of everyday perception. In understanding logical pictures, on the contrary, the individual requires specific cognitive schemata (so-called graphic schemata) in order to read off information from the visuo-spatial configuration (Lowe, 1996; Pinker, 1990).

Mental models, whether constructed during picture comprehension or during text comprehension, are internal depictive representations. They have inherent structural features in common with the depicted object. That is, they represent the object based on a structural or functional analogy (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). Such an analogy does not imply that mental models represent only spatial information. A mental model can represent, for example, also the increase or decrease of birth rates or incomes during a specific period of time (as it can be described in a text or displayed in a line graph), although birth rates and incomes are certainly not spatial information.

Although various findings indicate that mental model construction draws from capacity of the visuo–spatial part of working memory (Baddeley, 1992; Kruley, Sciama, & Glenberg, 1994; Sims & Hegarty, 1997), there are fundamental differences between mental models and visual images. First, since mental models are not bound to specific sensory modalities, they can be considered as more abstract than perceptual images. Second, mental models differ from visual images with regard to their informational content. On the one hand, a task-oriented selection takes place in mental model construction: Only those parts of the graphic configuration are included in the process of structure mapping which seem to be relevant for coping with present or anticipated tasks. On the other hand, the mental model is elaborated through information from world knowledge and, thus, also contains information which is not included in the picture.

As with text comprehension, new information can be read off from the model through processes of model inspection, which elaborates the propositional representation. In other words: There is a continuous interaction between the propositional representation and the mental model both in text comprehension and in picture comprehension. In text comprehension, the starting point of this interaction is a propositional representation which is used to construct a mental model. This model can again be used to read off new information in order to further elaborate the propositional representation. In picture comprehension, the starting point of the interaction is a mental model, which is used to read off new information that is also added to the propositional representation. Furthermore, as is indicated by the dotted diagonal arrows in Fig. 2, there is a possibility of an interaction between the text surface representation and the mental model as well as a possibility of an interaction between the perceptual representation of the picture and the propositional representation.

Accordingly, there is no one-to-one relationship between external and internal representations. On the one hand, a text as an external descriptive representation leads to both an internal descriptive and an internal depictive mental representation. A picture, on the other hand, as an external depictive representation leads to both an internal depictive and an internal descriptive mental representation.

Although constructing propositional representations and mental models could be considered at first glance as a kind of dual coding, our view is fundamentally different from the traditional dual coding theory. First, dual coding is assumed here not only for the processing of pictures, but also for the processing of words and texts. Second, the construction of a mental model is regarded as more than simply adding a further code that elaborates the mental representation and provides a quantitative advantage compared to a single code. Rather, the essential point is that propositional representations and mental models are based on different sign systems and different principles of representation, which complement one another.

4. Research questions and hypotheses

According to the model presented above picture comprehension is considered as a process of structure mapping between a graphic surface representation and a mental model representation. As there are often more possibilities to illustrate a subject matter, one has to decide which visualization seems to be best suited under the given circumstances. This decision depends also on the given task, because one picture may be more effective to solve a certain task whereas another picture is better suited to solve another task, even if the pictures contain the same information. Accordingly, the first research question investigated in this paper is whether the form of visualization influences the mental model structure. A further research question is how the process of mental model construction is influenced by adding pictures to a text. Is this always beneficial for learning or may picture comprehension even interfere with text comprehension?

To answer these questions, one has to consider that the informational content of a representation depends on its structure and on the procedures which operate on this structure. Two representations are informationally equivalent if all information that can be taken from one representation can also be taken from the other representation (Larkin & Simon, 1987). As a piece of information can be relevant for some tasks and irrelevant for other tasks, it is possible to define the informational content of a representation with respect to a specific set of tasks. Accordingly, two representations are (in a task-specific sense) informationally equivalent, if both allow the extraction of the same information which is required to solve the specific tasks. Nevertheless, extracting this information can be easier from one representation than from the other representation. Thus, despite their informational equivalence, the two representation can differ in their computational efficiency.

In Fig. 1 described above the carpet and circle diagrams are informationally equivalent with respect to *time difference tasks* and *circumnavigation tasks*. In the time difference tasks, the learner is asked to find time differences between specific cities, as for example in "*What time and which day is it in Los Angeles when it is Tuesday 2 o'clock p.m. in Tokyo?*". In the circumnavigation tasks the learner has to travel mentally around the world as, for example, in "*Why did Magellan's sailors believe that they arrived on a Wednesday after sailing around the world, although it was already Thursday?*". Both kinds of question can be answered with both forms of visualizations. Nevertheless, the two kinds of pictures differ in their computational efficiency with respect to the two different kinds of tasks. The *structure mapping hypothesis* assumes that structural features of the external graphic representation are mapped onto a mental model. Accordingly, different forms of informationally equivalent visualizations should result in mental models with different structures and, thus, with different computational efficiencies for different tasks.

The theoretical predictions of the *structure mapping hypothesis* are listed in Table

1. On the one hand, performance in solving time difference tasks with a mental model based on carpet diagrams should be higher than performance with a mental model constructed from circle diagrams. Accordingly, prediction 1 is: $\text{TimeDifference}_{\text{Carpet}} > \text{TimeDifference}_{\text{Circle}}$. On the other hand, performance in solving circumnavigation tasks should be higher with a mental model based on circle diagrams than with a model based on carpet diagrams. This results in prediction 2: $\text{Circumnavigation}_{\text{Carpet}} < \text{Circumnavigation}_{\text{Circle}}$.

Considering the second research question whether adding picture to a text supports or interferes with text comprehension three hypotheses are investigated: The dual coding hypothesis, the structure support hypothesis, and the structure interference hypothesis. The predictions derived from these hypothesis with regard to time difference tasks performance and circumnavigation tasks performance are also included in Table 1.

According to the dual coding theory (Paivio, 1986) texts combined with pictures lead to more elaborated cognitive structures than texts without pictures. Consequently, learning from text with pictures leads to better memory for the learned information and to better performance in knowledge acquisition. As the dual coding theory makes no assumptions whether one kind of picture is better than another one, the *dual coding hypthesis* predicts that learners who received a text with carpet diagrams or with circle diagrams will perform better in solving time difference tasks than learners who received the text without pictures. Accordingly, prediction 3 is: $\text{TimeDifference}_{\text{Carpet}} > \text{TimeDifference}_{\text{TextOnly}}$, and prediction 4 is: $\text{TimeDifference}_{\text{Circle}} > \text{TimeDifference}_{\text{TextOnly}}$. Similarly, it predicts that learners who received a text with carpet diagrams or with circle diagrams will perform better in solving circumnavigation tasks than learners who received a non-illustrated text. This results in prediction 5: $\text{Circumnavigation}_{\text{Carpet}} > \text{Circumnavigation}_{\text{TextOnly}}$, and prediction 6: $\text{Circumnavigation}_{\text{Circle}} > \text{Circumnavigation}_{\text{TextOnly}}$.

If the form of visualization affects the structure of mental models, however, the dual coding assumptions could be too simplistic. Instead, it is possible that pictures with task-appropriate visualization support the mental model construction whereas pictures with task-inappropriate visualization may interfere with mental model con-

Table 1
Theoretical predictions regarding the structure mapping, the dual coding, the structure support and the structure interference hypothesis

	Time difference task performance	Circumnavigation task performance
Structure mapping	p1: Ca>Ci	p2: Ca<Ci
Dual coding	p3: Ca>To	p5: Ca>To
	p4: Ci>To	p6: Ci>To
Structure support	p7: Ca>To	p8: Ci>To
Structure interference	p9: Ci<To	p10: Ca<To

p=prediction, To=Text only, Ca=Text with carpet diagrams, Ci=Text with circle diagrams.

struction. These two possibilities are referred to as the structure support hypothesis and the structure interference hypothesis.

Previous research has suggested that pictures support mental model construction if learners have poor learning prerequisites (Levie & Lentz, 1982; Mayer, 1993), although Kirby (1993) found that pictures can also have non-beneficial effects on learning from text. If picture comprehension includes a process of structure mapping between the picture and the mental model, then one can assume that adding task-appropriate pictures facilitates the construction of a task-appropriate mental model. Accordingly, the *structure support hypothesis* predicts that adding carpet diagrams to the text helps learners to construct a mental model appropriate for solving time difference tasks. This results in prediction 7: $\text{TimeDifference}_{\text{Carpet}} > \text{TimeDifference}_{\text{TextOnly}}$. Similarly, the addition of circle diagrams to a text should help learners to construct a mental model appropriate for solving circumnavigation tasks: Accordingly, prediction 8 is: $\text{Circumnavigation}_{\text{Circle}} > \text{Circumnavigation}_{\text{TextOnly}}$. Thus, one can predict that performance would in both cases be better than after learning from non-illustrated text.

On the contrary, if a picture with a task-inappropriate form of visualization is added to the text, then learners could be hindered in their comprehension process, because the structure of the picture interferes with the required mental model construction. For example, circle diagrams would interfere with the construction of a carpet-like model, and carpet diagrams would interfere with the construction of a circle-like model. Accordingly, the *structure interference hypothesis* predicts that individuals after learning from a text with circle diagrams would perform poorer in solving time difference tasks than after learning from a text without diagrams. Accordingly, prediction 9 is: $\text{TimeDifference}_{\text{Circle}} < \text{TimeDifference}_{\text{TextOnly}}$. Similarly, one would expect that individuals after learning from a text with carpet diagrams would show lower performance in solving circumnavigation tasks than after learning from non-illustrated text. This results in prediction 10: $\text{Circumnavigation}_{\text{Carpet}} < \text{Circumnavigation}_{\text{TextOnly}}$.

To summarize, the dual coding hypothesis and the structure mapping hypothesis make different assumptions concerning an interaction between kind of presentation and kind of tasks. The dual coding hypothesis predicts that both kinds of pictures (carpet and circle pictures) would enhance performance for both kinds of tasks (time difference and circumnavigation tasks) and, accordingly, does not assume an interaction. The structure mapping hypothesis, on the contrary, predicts different effects of different pictures on performance in different tasks and, thus, assumes a specific interaction between the kind of presentation and the kind of tasks.

5. Method

In order to test these hypotheses an experiment was conducted according to a 3x2 factor design including the between groups factor *kind of presentation* (text with carpet diagrams, text with circle diagrams, text only) and the within groups factor *kind of task* (time difference tasks, circumnavigation tasks). Sixty university students majoring in different fields participated in the experiment. Subjects were randomly assigned to three groups of 20 persons each corresponding to three experimental conditions (text with carpet diagrams, text with circle diagrams, text only). Learning material was a hypermedium about time differences on earth consisting of 22 text cards with a total of 2750 words which was presented on a computer screen. The text-only group received only this hypertext. The carpet group received the hypertext with carpet diagrams (see top of Fig. 1), and the circle group was given the hypertext with circle diagrams (see bottom of Fig. 1). Texts in the three experimental treatments were informationally equivalent except for some differences in wording which were necessary in order to refer to the different figures. Likewise, the carpet diagrams and circle diagrams were informationally equivalent with regard to the time difference and circumnavigation questions. The diagrams were presented together with the corresponding text paragraphs. Presentation rate of the learning material was controlled by the participants.

Subjects' prior knowledge was assessed in a pre-test by an essay type test concerning time differences on earth. Furthermore, subjects' verbal and spatial intelligence were assessed by using subtests of the Intelligenzstrukturtest IST 70 (Amthauer, 1973). The experimental session consisted of three phases: practice, learning and test. In the *practice phase*, the subjects were given a hypermedium on a different topic in order to learn how to handle the medium. In the *learning phase* subjects acquired knowledge about the reasons for time differences on earth. In the *test phase*, the subjects were given a comprehension test consisting of time difference tasks (16 items) and circumnavigation tasks (16 items). Items were presented in multiple choice format on a computer screen and had to be answered without consulting the instructional material. For each subject time difference task performance was determined as the number of correctly answered time difference tasks. Similarly, circumnavigation task performance was determined as the number of correctly answered circumnavigation tasks. Time difference task and circumnavigation task performance defined different levels of the factor *kind of task* with repeated measures. The experiment was conducted in individual sessions and lasted approximately two hours.

6. Results

The three experimental groups were homogenous with regard to prior knowledge, verbal ability, and spatial ability. There were also no significant group differences with respect to learning time.

Means and standard deviations of the time difference tasks performance and the circumnavigation tasks performance within the three groups are shown in Table 2. With regard to the time difference tasks the best performance was shown by the text-only group with a mean of 10.30 correctly solved items, closely followed by the carpet group with a mean of 9.05 correctly solved items. The poorest performance was exhibited by the circle group, where the subjects answered on the average only 7.20 items correctly. With regard to the circumnavigation tasks, on the contrary, the

Table 2
Time difference tasks and circumnavigation tasks performance

		Carpet group (n=20)	Circle group (n=20)	Text-only group (n=20)
Time difference	*M*	9.05	7.20	10.30
tasks performance	*SD*	2.86	2.17	3.34
Circumnavigation	*M*	8.50	10.85	10.15
tasks performance	*SD*	2.56	3.23	3.12

circle group showed the highest performance. The circle group solved an average of 10.85 items correctly, followed by the text-only group with an average of 10.15 correctly solved items, whereas the carpet group answered only 8.50 items successfully.

An ANOVA with the between groups factor *kind of presentation* (text+carpet/text+circle/text-only) and the within groups factor *kind of task* (time difference tasks, circumnavigation tasks) revealed no significant main effect for *kind of presentation* ($F(2.57)=2.16$; $p=0.124$), but a significant main effect for *kind of task* ($F(2.57)=5.00$; $p=0.029$) as well as a highly significant interaction effect *kind of presentation*×*kind of task* ($F(2.57)=9.25$; $p<0.001$). The highly significant interaction effect supports the structure mapping hypothesis. Although we had predicted specific prior contrasts based on different hypotheses, we will use in the following only post hoc comparisons in order to test these predictions in a conservative way.

Duncan's post hoc comparison of the group means revealed that the carpet group performed marginally better than the circle group with the time difference tasks ($p=0.102$), but performed significantly worse than the circle group with the circumnavigation tasks ($p=0.049$). These results confirm predictions 1 and 2 of the structure mapping hypothesis. On the contrary, no empirical support was found for the dual coding hypothesis according to Duncan's post hoc test. Whereas predictions 3, 4 and 5 were falsified by differences in the opposite direction, prediction 6 was also not confirmed by a significant group difference ($p=0.537$). Accordingly, the dual coding hypothesis seems not to be a good candidate to explain the empirical findings.

There was also no evidence found for the *structure support hypothesis*, as none of its predictions was confirmed by significant group differences. Prediction 7 was falsified by a tendency in the opposite direction. Prediction 8 was also not supported by a significant result ($p=0.537$). The empirical data fit relatively well, however, to the *structure interference hypothesis*. On the one hand, circle diagrams interfered with the construction of a mental carpet model, as prediction 9 was confirmed by a highly significant difference ($p=0.009$). On the other hand, the carpet diagrams seem to have interfered with the construction of a mental circle model. A difference was found according to prediction 10, although it failed significance due to the high intra group variance (p=0.145).

7. Discussion

Text comprehension and picture comprehension are goal-oriented processes of the human cognitive system, in which the individual actively selects and processes verbal as well as pictorial information in order to construct mental representations that seem to be suited to cope with present or anticipated demands. Thus, a comprehensive theory of text and picture comprehension should take this active and constructive nature of information processing into account and should be embedded into a broader framework of human cognition. Such a theory should take into account that pictures can visualize one and the same subject matter in different ways. It should be able to predict what effects different forms of visualization will have on the structure of the mental representation created in the comprehension process as well as on the computational efficiency of this representation for specific tasks. The theory should also be able to predict under which conditions the addition of a picture to a text is beneficial for comprehension or knowledge acquisition and under which conditions it has rather a detrimental effect.

The findings of the study presented above indicate that the dual coding theory is not a satisfactory basis for the development of such a theory. The dual coding theory does not take into account that a subject matter can be visualized in different ways and that the form of visualization affects the structure of the mental representation. Furthermore, it assumes that adding pictures to a text is generally beneficial for learning, that is, it neglects that pictures can also have negative effects because a picture may interfere with mental model construction.

The combined model of text and picture comprehension presented above seems to be better suited for an explanation of the empirical findings of our study. The model considers picture comprehension as a process of analogical structure mapping. Accordingly, the model allows us to explain why the form of visualization used in a picture affects the structure of the mental model created during picture comprehension. Obviously, the surface structure of the picture is mapped (at least partially) onto the structure of the mental model and, thus, affects the computational efficiency of this model for specific tasks.

The model allows us to explain why adding pictures to a text is not always beneficial, but can also have detrimental effects on the construction of task-appropriate mental representations as has been demonstrated by Kirby (1993). Previous research on children's processing of narrative texts has shown that the poor readers profit generally more from text illustrations with regard to comprehension and learning than good readers (Cooney & Swanson, 1987; Levie & Lentz, 1982; Rusted & Coltheart, 1979; Mastropieri & Scruggs, 1989). This suggests that poor readers are able to construct a mental model from a text with pictures, whereas they would fail on the basis of a text alone. Similar results have been found for adult learners' processing of expository texts. Learners with low prior knowledge benefit from pictures in a text, whereas learners with higher prior knowledge seem to be able to construct a mental model of the described content also only from the text (Mayer, 1997). These results and our own findings suggest that pictures facilitate learning only if individuals have low prior knowledge and if the subject matter is visualized in a task-appropriate way. However, if good readers with high prior knowledge receive a text with pictures in which the subject matter is visualized in a task-inappropriate way, then these pictures may interfere with the construction of a task-appropriate mental model.

From the perspective of practice, the findings of our study emphasize that in the design of instructional material including texts and pictures the form of visualization used in the pictures should be considered very carefully. The question is not only which information is to be conveyed. One must also ask whether the form of visualization used in the picture supports the construction of a task-appropriate mental model. Good graphic design is not only important for individuals with low prior knowledge who need pictorial support in constructing mental models. Well-designed pictures are also important for individuals with high prior knowledge because these individuals can be hindered in their mental model construction through inappropriate forms of visualization.

The model of text and picture comprehension presented above provides a theoretical framework for research on the construction of mental representations through the combined processing of verbal and pictorial information. The empirical results of our study can be explained within this framework and, at the same time, they help to specify the relationships hypothesized in this framework. In order to make the model fruitful both under theoretical and practical aspects, however, further research is needed. Future studies should focus especially on the interplay between external and internal forms of representation during the acquisition of knowledge under different conditions which include different learning domains, groups of subjects, kinds of learning requirements, different text characteristics and different forms of visualization. The findings of these studies should help to further elaborate an integrative view of text and picture comprehension.

Acknowledgements

The authors thank Beatriz Barquero, Susan Goldman, John Kirby, Ric Lowe, Richard Mayer, Phil Moore, Peter Reimann and Tina Seufert for helpful discussions about theoretical issues of this paper.

References

Amthauer, R. (1973). Intelligenz-Struktur-Test 70 (IST 70). Göttingen Hogrefe.

Baddeley, A. (1992). Working memory. *Science*, *255*, 556–559.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293–332.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*, 149–210.

Cooney, J. B., & Swanson, H. L. (1987). Memory and learning disabilities: An overview. In H. L. Swanson (Ed.), *Memory and learning disabilities: Advances in learning and behavioral disabilities* (pp. 1–40). Greenwich, CT: JAI.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: algorithm and examples. *Artificial Intelligence*, *41*, 1–63.

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 197–241). Cambridge: Cambridge University Press.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Kirby, J. (1993). Collaborative and competitive effects of verbal and spatial processes. *Learning and Instruction*, *3*, 201–214.

Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.

Kruley, P., Sciama, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory and Cognition*, *22*, 261–272.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65–99.

Levie, H. W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal*, *30*, 195–232.

Levin, J. R., Anglin, G. J. & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton, (Eds.), The psychology of illustration, Vol. 1 (pp. 51-86). New York: Springer.

Lowe, R. K. (1996). Background knowledge and the construction of a situational representation from a diagram. *European Journal of Psychology of Education*, *11*, 377–397.

Mastropieri, M. A., & Scruggs, T. E. (1989). Constructing more meaningful relationships: Mnemonic instruction for special populations. *Educational Psychology Review*, *1*, 83–111.

Mayer, R. E. (1993). Illustration that instruct. In R. Glaser (Ed.), Advances in instructional psychology, Vol. 5 (pp.253-284). Hillsdale, NJ: Erlbaum.

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

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

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

Peirce, C. S. (1906). Prolegomena to an apology for pragmaticism. *The Monist*, 492-546.

Pinker, S. (1990). A theory of graph comprehension. In R. Frele (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Erlbaum.

Rusted, J. M., & Coltheart, M. (1979). Facilitation of children's prose recall by the presence of pictures. *Memory and Cognition*, *7*, 354–359.

Schnotz, W. (1993). On the relation between dual coding and mental models in graphics comprehension. *Learning and Instruction*, *3*, 247–249.

Schnotz, W. (1994). *Aufbau von Wissensstrukturen*. Weinheim: Beltz.

Shepard, R. N. (1984). Ecological constraints on internal representations: Resonant kinematics of perceiving, thinking, and dreaming. *Psychological Review*, *91*, 417–447.

Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-tasks studies. *Memory and Cognition*, *25*, 321–332.

Ullmann, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

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

van Someren, M. W., Reimann, P., Boshuizen, H. P. A., & de Jong, T. (Eds.). (1998). *Learning with multiple representations*. Oxford: Elsevier.

Weaver III, C. A., Mannes, S., & Fletcher, C. R. (Eds.). (1995). *Discourse comprehension*. Hillsdale, NJ: Erlbaum.

Wertheimer, M. (1938). *Laws of organization in perceptual forms in a source book for Gestalt Psychology*. London: Routledge and Kegan Paul.

Winn, W. D. (1994). Contributions of perceptual and cognitive processes to the comprehension of graphics. In W. Schnotz, & R. Kulhavy (Eds.), *Comprehension of graphics* (pp. 3–27). Amsterdam: Elsevier.