Three Turtles in Danger: Spontaneous Construction of Causally Relevant Spatial Situation Models

Georg Jahn
University of Regensburg

In 4 experiments, the author explored the spontaneous construction of spatial situation models during discourse comprehension by using the sentence-recognition paradigm of J. D. Bransford, J. R. Barclay, and J. J. Franks (1972). In Experiment 1, signaling causal relevance of spatial relations was a necessary precondition for replicating their original finding of spontaneously constructed spatial representations. Causal relevance was ensured in the subsequent experiments by a judgment task indirectly demanding the evaluation of described spatial relations with regard to causal relevance. Participants spontaneously constructed spatial situation models of text presented auditorily or visually. Effects of spontaneous construction were more reliable when encoding was easier. The results suggest a revised interpretation of J. D. Bransford et al.’s study and corroborate recent evidence showing that relevant spatial information in texts is reliably represented.

Narrative texts such as fairy tales, stories, and novels often invite readers and listeners to imagine narrated events (Gerrig, 1993). In a story by P. G. Wodehouse (1930/2000), the main character sets up a trap in an office building and invites the reader to follow the spatial description he gives afterward:

The first thing you need in matters of this kind, as every general knows, is a thorough knowledge of the terrain. Not know the terrain, and where are you? [. . .] I won’t draw a plan, because my experience is that, when you’re reading one of those detective stories and come to the bit where the author draws a plan of the Manor, showing room where body was found, stairs leading to passage-way, and all the rest of it, one just skips. I’ll simply explain in a few brief words. (p. 38)

Without some idea of the location of the trap and the layout of rooms in the building, readers would have difficulties comprehending the events in the rest of the story, in which the trap plays an important role. However, spatial information without noticeable relevance for unfolding events can fail to engage readers, as Wodehouse (1930/2000) notes.

From the common experience of imagery during reading and the importance of spatial information for understanding events, one can infer that recipients represent spatial configurations and track explicit or implied movements while comprehending narrated events. Surprisingly, in a review of research on narrative comprehension, Zwaan and Radvansky (1998) stated as a summary of experimental findings that “. . . there currently is no strong evidence that readers spontaneously track spatial information during comprehension. However, they are able to do so when asked” (p. 168). In this article, I explore why common experience and experimental findings currently drift apart on the issue of space in narrative comprehension. The main argument is that space lacks relevance in most experimental texts and therefore has not been spontaneously represented by participants in the respective studies. Four replications of a classic study (Bransford, Barclay, & Franks, 1972) that is famous for the rare demonstration of spontaneously constructed spatial representations are reported. They suggest that causal relevance of space was a crucial precondition for the original finding. In addition to causal relevance of space, effects of presentation modality and ease of encoding are systematically explored.

Spatial representations constructed during comprehension belong to the situation-model level in current theories of discourse comprehension. Comprehending discourse is an active process limited in resources, during which multilevel representations of text are constructed. Reading and listening to narratives not only gives rise to representations of the text itself, its wording, or its propositional structure (i.e., surface form and textbase, van Dijk & Kintsch, 1983), but readers and listeners, trying to comprehend, follow the text as a set of instructions to construct representations of what the text is about (Garnham, 1996; Gernsbacher, 1990). These mental models (Craik, 1943/1967; Johnson-Laird, 1983), which are elaborated by recourse to world knowledge, are called situation models in text-comprehension theories (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). Situation models are hypothesized to represent protagonists and objects mentioned in narrative texts, along with the intentions and emotions of protagonists and events. The representations of events in situation models include causal, temporal, and spatial relations.

Georg Jahn, Department of Psychology, University of Regensburg, Regensburg, Germany.

I thank Sylvia Grassl and Andreas Struck for their help in conducting the experiments. I am also grateful to Gabriel Radvansky, Michael Masson, Phil Johnson-Laird, Till Pfeiffer, and Johannes Hönekopp for their helpful comments on drafts of this article. Two of the reported experiments were part of Georg Jahn’s doctoral dissertation conducted at the University of Regensburg, and two experiments were completed at the Chemnitz University of Technology, Chemnitz, Germany.

Correspondence concerning this article should be addressed to Georg Jahn, who is now at the Center for Cognitive Science, Institute of Computer Science and Social Research, University of Freiburg, Friedrichstr. 50, D-79098 Freiburg, Germany. E-mail: georg.jahn@cognition.uni-freiburg.de
Spatial situation models are usually postulated to include visuo-spatial representations (e.g., Fincher-Kiefer, 2001; Glenberg, 1997; Langston, Kramer, & Glenberg, 1998; Zwaan, 1999). Under certain circumstances, even a single sentence may initiate an elaborate situation model. However, readers and listeners usually have to strategically select what to include, which cues to follow, and which inferences to integrate while constructing representations during text comprehension (Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992).

Causal (Long, Golding, & Graesser, 1992; Trabasso & Magliano, 1996) and temporal relations (Rinck, Hähnel, & Becker, 2001; Zwaan, 1996) are consistently found to be spontaneously tracked by comprehenders. However, it has proved difficult to demonstrate spontaneous tracking of spatial relations in studies of text comprehension. In many experiments, comprehenders obviously avoided the effort of constructing spatial representations (de Vega, 1995; Levine & Klin, 2001; Zwaan, Radvansky, Hilliard, & Curiel, 1998; Zwaan & van Oostendorp, 1993). Several important preconditions for the spontaneous construction of spatial situation models were often flawed in previous studies but are usually fulfilled in naturalistic discourse.

Preconditions of Spontaneous Spatial Situation Model Construction

The effective preconditions may be summarized by two main conditions. First, constructing the spatial representation from the text must not be too difficult. Second, tracking spatial relations must seem advantageous to comprehenders; that is, it must seem to be worth the effort.

The first main condition, the difficulty of constructing a spatial representation from text, depends first of all on the number of entities that have to be represented (Morra, 2001). Furthermore, the structure of spatial descriptions affects the construction. In studies of spatial mental models, descriptions were easier to transform into mental models when they were determinate (Baguley & Payne, 2000; Mani & Johnson-Laird, 1982), continuous (Denis & Cocude, 1992; Ehrlich & Johnson-Laird, 1982; Oakhill & Johnson-Laird, 1984), and condensed (Zwaan & van Oostendorp, 1994).

In naturalistic discourse, no more than four entities are usually relevant at a given moment during comprehension. Additionally, the structure of spatial descriptions usually follows conventions (Talmy, 1980) that support comprehending and transforming descriptions (e.g., Levelt, 1996; Plumert, Spalding, & Nichols-Whitehead, 2001). For example, spatial descriptions are conventionally centered on a salient and static reference object. Moreover, prior knowledge about unique or usual schematic spatial layouts (Dutke, 1996) eases the transformation of descriptions. And finally, the cognitive abilities of the individual comprehender affect the ease of model construction (e.g., Denis & Zimmer, 1992; Giesen & Preece, 1984).

If the construction of a spatial situation model is not too difficult, the second main precondition for spontaneous construction becomes crucial—the expectation that tracking spatial relations will be worth the effort. The motivation to track spatial relations is mainly determined by the goals of the comprehender (Graesser & Kreuz, 1993).

Usually the goal of the comprehender may be described as a general effort to understand the meaning an author attempts to convey (Foertsch & Gernsbacher, 1994). Therefore, the expected and inferred intentions of the author affect the probability of spatial representation (Morrow, Bower, & Greenspan, 1989). Mediated by the motivation to follow the author’s intentions, spatial representations become more likely if the author signals their importance. This is often accomplished implicitly by foregrounding space via textual cues (e.g., de Vega, 1995).

Related to author intentions, the text genre influences comprehenders’ expectations about the importance of spatial relations (Zwaan, 1994); for example, there are text genres, such as hiking guides or soccer broadcasts, that directly focus on space (Schirra, 1995). In narrative texts, the focus is on causality (Trabasso & Magliano, 1996), and a motivated comprehender will try to keep up with causal developments. It is this intention to track causality that turns causal relevance into an effective signal to represent space in the comprehension of narratives, as is discussed in more detail in the Causal Relevance section. The motivation to construct causally relevant spatial representations is presumably especially high if events are still indeterminate and unfolding and if the author leaves the comprehender something to fill in and to imagine (Duffy, Shinjo, & Myers, 1990; McNamara, Kintsch, Songer, & Kintsch, 1996). Otherwise, even causally relevant spatial descriptions may fail to stimulate spontaneous construction activity.

All these preconditions concerning the ease of construction, pragmatics, and motivation must be taken into consideration if experimental results are to be generalized into naturalistic discourse comprehension. It is not sufficient to present spatial information embedded in a story if participants have no reason to integrate it in a situation model. In naturalistic discourse, spatial information usually bears relevance. Growing evidence points to causal relevance of space as a crucial precondition for spontaneous spatial representation during the comprehension of narrative texts.

Causal Relevance

Although often studied in isolation, causality and space may be closely intertwined in narratives (Zwaan & Radvansky, 1998; Zwaan & van Oostendorp, 1994). For example, if a protagonist wants to keep his or her actions secret, the location of the other protagonists and the spatial layout in which the events take place are central to the comprehension of events. The idea behind causal relevance is that the effort to comprehend the causal structure of events extends to an effort to represent spatial relations if there is such a strong link between the causal chain of events and spatial relations.

Three studies that focused on causally relevant spatial relations yielded equivocal results. Taken together, they suggest that construction activity has to be spontaneous to reveal effects of causal relevance. One of the studies was conducted by Rice (1999), who used a short story about an episode in the Civil War. In this study, participants more accurately verified spatial relations when these were causally relevant.

In a second study in which they too used verification of spatial relations, Friedman and Miyake (2000) found no effects of causal relevance, although their texts were suitable. They give as an example for their texts a story about a director of a museum who thinks about where to place surveillance cameras to keep the whole
exhibition in view. Friedman and Miyake interpret their negative finding as evidence that causality and space do not interact during text comprehension. This is probably only true under their experimental conditions, because they did not study spontaneous construction of spatial representations but strongly induced representation of space as a mental map with memory instructions and control tasks and by providing a floor plan. The tasks and instructions that directly guided participants to attend to spatial descriptions could easily have prevented causal relevance from making a difference.

In a third study, Sundermeier, van den Broek, and Zwaan (2004; Zwaan & Radvansky, 1998) focused on spontaneous tracking of causally relevant spatial relations. Causal relevance of object locations varied across versions of stories. During story versions that described objects to be in locations where they might cause events, the object names were answered faster in a probe recognition test. This result and think-aloud protocols indicated that comprehenders anticipated events; for example, they anticipated that a thumbtack on the floor would be causally relevant, because somebody walking around barefoot is in danger of getting hurt.

Causal relevance has to be detected by comprehenders before it can foster the construction of spatial situation models. If objects or protagonists take typical roles in a scenario, or if protagonists’ goals map onto roles in a scenario, comprehenders will make use of their scenario knowledge and will anticipate the following events in a story. This use of prior knowledge in comprehension is central to Sanford and Garrod’s (1998) scenario-mapping model (see also Schmalhofer, McDaniel, & Keefe, 2002).

However, the knowledge on which the detection of causal relevance is based does not have to be of a full scenario, script, or schema (Hannigan & Reinitz, 2001). Glenberg and Robertson’s (1999) indexical hypothesis explains the detection of causal relevance on the basis of more elementary knowledge for at least concrete events. The indexical hypothesis postulates an embodied representation of meaning (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Clark, 1999), which develops from the knowledge about the affordances that objects mentioned in a story have for protagonists. For example, the thumbtack affords danger for a barefoot protagonist if it is lying on the floor, and a bush affords camouflage for a soldier hiding from an enemy. Spatial representations, according to the indexical hypothesis, are built from perceptual symbols (Barsalou, 1999; Stanfield & Zwaan, 2001).

Whether based on scenario knowledge or on a new combination of more elementary knowledge, causal relevance may induce the spontaneous construction of spatial situation models, because the recipient forms expectations about predictable events in a story (e.g., Chambers et al., 2002; Fincher-Kiefer, 1993; Johnson, Bransford, & Solomon, 1973; Klin, Guzmán, & Levine, 1999; Long et al., 1992). Alternatively, if a narrated event was not expected, causally relevant spatial relations may be represented because the recipient tries to construct an explanation for the event (Trabasso & Suh, 1993). In several studies that explored similar ideas, causal relevance proved to be crucial for spontaneous spatial representation under the labels functionality of spatial relations (Radvansky & Copeland, 2000; Radvansky, Copeland, & Zwaan, 2003), relevance for protagonist’s goals (Levine & Klin, 2001; Morrow, Bower, & Greenspan, 1990), or required for comprehension (Glenberg & McDaniel, 1992; Hakala, 1999).

There is one study on spontaneous construction of spatial representations from text that usually is not regarded as showing the effect of causal relevance. It is the prominent illustration of situation model construction by Bransford et al. (1972, Experiment 1), who used a sentence-recognition paradigm (see also Garnham, 1981; Glenberg & McDaniel, 1992; Hakala, 1999; Radvansky & Copeland, 2001; Reyna & Kiernan, 1994; Rinck et al., 2001). Participants were instructed to listen carefully to sentences like 1A and 2A below and were told that they would be asked questions about these sentences afterward. Then after a pause of 3 min, the participants were presented with the sentences heard and with new sentences like 1B and 2B in a recognition test.

1A. Three turtles rested on a floating log, and a fish swim beneath them. 1B. Three turtles rested on a floating log, and a fish swim beneath it. 2A. Three turtles rested beside a floating log, and a fish swim beneath them. 2B. Three turtles rested beside a floating log, and a fish swim beneath it.

The new Sentence 1B was often falsely recognized as the old Sentence 1A by participants, whereas participants who heard 2A could better identify Sentence 2B as new. This difference between the discriminability of sentence pairs 1A–1B and 2A–2B cannot be explained by surface structure differences, because in both pairs the sentences differ only in the last word, the pronoun, which is them or it. Usually the higher discriminability of sentences like 2A and 2B is explained by the spatial situation models to which the sentences give rise (e.g., Glenberg, Meyer, & Lindem, 1987). Sentences 1A and 1B describe spatially similar situations, the turtles on the log and the fish beneath both, and therefore the spatial relations in the situation models participants built from Sentences 1A and 1B are the same. Hence, Sentences 1A and 1B can be called a same-model pair. In contrast, Sentences 2A and 2B describe spatially different situations. Sentence 2A and Sentence 2B both place the turtles horizontally beside the log, but Sentence 2A places the fish beneath the turtles, and Sentence 2B places the fish beneath the log. Consequently, the spatial relations in the situation models of Sentences 2A and 2B differ, and Sentences 2A and 2B are an example of a different-models pair.

The spatial relations described in the sentences refer to relative locations of entities in a scene rather than to the location of the whole scene. A comparison with the material used by Garnham (1981) for the same recognition paradigm illustrates this difference. One of Garnham’s “confusable” sentence pairs was “The judge got his contact lenses (from the optician/in the optician’s).” The corresponding nonconfusable pair was “The judge answered a telephone call (from the optician/in the optician’s)” (p. 565). The sentences in Garnham’s nonconfusable pairs differ with regard to the implied or mentioned location and also with regard to implied or mentioned protagonists. They were easier to discern in the recognition test than were confusable sentences after a study phase in which participants were asked “to rate how easily they could imagine the situation described by each sentence” (p. 562).

The model type effect on discriminability found by Bransford et al. (1972) depends on the construction of spatial situation models. Without a specific instruction, participants constructed spatial situation models during the study phase and retrieved and compared them to spatial situation models constructed from sentences in the recognition test. Therefore, this study is regarded as demonstrating spontaneous construction of spatial situation models in an exper-
models would have improved the recognition performance for different-models sentences but not for same-model sentences (the model-type effect).

In addition to causal relevance, I investigated presentation modality and ease of encoding with regard to possible effects on spontaneous construction and recognition performance in Experiments 2–4 to test an explanation of difficulties in replicating Bransford et al.’s (1972) study in terms of presentation modality. In Experiments 2 and 3, I used double auditory presentation; in Experiment 4, I used self-paced visual presentation.

Experiment 1: Replication Without Danger Judgment

In the original study by Bransford et al. (1972) as discussed in the Causal Relevance section, a strong model-type effect on recognition performance was found. Old and new different-models sentences were discriminable in the recognition test, whereas old and new same-model sentences were indiscriminable. Two of three examples given for the original material described predator scenes and, therefore, fulfilled causal relevance. If causal relevance increases the probability of spontaneous model construction, as I assume on the basis of theoretical considerations and some empirical evidence, a majority of predator scenes in the material could have been a precondition for the original finding, which is usually associated with a peaceful scene, the famous turtles-and-fish example. In Experiment 1, scene type is systematically varied in addition to model type to test this revised interpretation.

Predator-scene and peaceful-scene, same-model and different-models sentences were constructed. They all described spatial layouts with a single object and two animals or groups of animals as in the examples given by Bransford et al. (1972). Participants heard sentences from same-model and different-models pairs in the study phase and were instructed, as in the original study, simply to listen carefully, because questions on the sentences would follow. After a delay of 3 min, old sentences and their same-model and different-models counterparts were presented one sentence at a time in a recognition test. The model type effect was predicted only for predator-scene sentences but not for peaceful-scene sentences, because causal relevance of space was signaled only with a predator animal present. Thus, a Scene Type $\times$ Model Type interaction was expected in the recognition test.

Method

Participants. Thirty-two psychology students from the University of Regensburg, Regensburg, Germany participated in the experiment to fulfill a curricular requirement.

Materials. A total of 96 experimental sentences and 12 filler sentences were constructed for the experiment. The sentence material was in German and is listed, with English translations, in the Appendix. The experimental sentences were built to resemble Bransford et al.’s (1972) materials. Each described two activities of animals located near an inanimate reference
object. For example, one sentence was: Two zebras graze next to a shrub and a lion trots towards it ( predator-scene same-model Sentence A). The pronoun it refers to shrub, so both activities, grazing and trotting, are located with reference to the shrub. By exchanging the pronoun, the second activity is made to refer to the zebras: Two zebras graze next to a shrub and a lion trots towards them ( predator-scene different-models Sentence B). Because the zebras and the shrub are described to be close together, exchanging the pronoun does not change the spatial relations in the scene. Therefore, the spatial situation model was expected to be the same for these two sentences that together constitute a same-model pair.

Different-models pairs of sentences that were parallel to each same-model pair were constructed. These described similar scenes with the same animals and the same inanimate object in a similar syntactical form, but the spatial relations differed between these scenes. One sentence of the different-models pair was Two zebras move away from a shrub and a lion trots towards it ( predator-scene different-models Sentence A). In the second sentence, again, only the pronoun it was replaced by them: Two zebras move away from a shrub and a lion trots towards them ( predator-scene different-models Sentence B). Now the zebras are described to be away from the shrub. Exchanging the pronoun changed the spatial relations and induced a different spatial situation model.

Experimental sentences that mentioned predator animals described predator scenes. Same-model and different-models sentences describing peaceful scenes were constructed by replacing the predator animal by a peaceful animal in all sentences. The peaceful animal for the given example was an antelope: Two zebras move away from a shrub and an antelope trots towards them (peaceful-scene different-models Sentence B; see the Appendix).

In total, 12 groups of experimental sentences were constructed, each consisting of same-model and different-models pairs describing predator scenes and same-model and different-models pairs describing peaceful scenes. Different kinds of spatial configurations were used. The spatial relations in some sentence groups were horizontal distances and directions involving a small (bowl, bush) or flat (puddle, hollow) reference object (see Sentence Groups 1–4 in the Appendix). Other sentence groups contained vertical spatial relations like the turtles-and-fish and robins-and-hawk examples of Bransford et al. (1972, Sentence Groups 5–8). Furthermore, there were configurations of animals surrounding others or surrounding an object as in the raccoons-and-dogs example of Bransford et al. (Sentence Groups 9 and 10) and configurations in which the reference object was an enclosed space with a defined front, for example, in front of a burrow (Sentence Groups 11 and 12). The sentences were written to sound fine in German. Therefore, it was not possible to keep the syntactical form constant across experimental sentence groups, but it was kept constant within each sentence group.

In addition to the 12 groups of experimental sentences, 6 pairs of filler sentences were constructed (see the Appendix). The syntactical form of the filler sentences was less restricted. They also described activities of animals and spatial relations, and they all mentioned two animals or groups of animals, for example (Filler Pair F1): A lynx is standing behind a horse as it kicks out. A lynx is running past behind a horse as it kicks out. Filler sentences in one pair differed in more than just a single pronoun, and therefore they should be easier to distinguish than experimental sentences of one pair.

During the study phase of the experiment, 18 sentences were presented to each participant, consisting of 1 sentence of each of the 12 groups of experimental sentences and 1 from each of the 6 pairs of filler sentences. Half of the experimental sentences presented to each participant were same-model sentences (3 describing predator scenes and 3 describing peaceful scenes), and half were different-models sentences (3 describing predator scenes and 3 describing peaceful scenes). In addition, balanced within the experimental sentences was the reference of pronouns presented to a single participant. In half of the sentences, the pronoun referred to the inanimate object; in the remaining sentences, the pronoun referred to the other animal. Between subjects, the experimental sentences in the study phase were counterbalanced with regard to sentence group combinations and kinds of scenes ( predator or peaceful), combinations of sentence groups and kinds of model pair (same model or different models), and combinations of sentence groups and kinds of pronoun reference.

The experimental sentences in the study phase were presented in random order. Filler sentences were inserted pseudorandomly. For each participant, the study phase began with a filler sentence. The remaining filler sentences were inserted with the restriction that no more than two filler sentences were presented successively. In this way, I ensured that filler sentences actually broke up the sequence of syntactically homogeneous experimental sentences.

During the recognition test of the experiment, 30 sentences were presented to each participant. Among those were the 12 experimental sentences from the study set (the old experimental sentences) and their 12 same-model or different-models counterparts (the new experimental sentences). Filler sentences in the recognition test were 3 old filler sentences from the participant’s study set and their 3 new counterparts. Again, there were restrictions on the order of presentation. Sentences in a same-model or different-models old–new pair are very similar. Therefore, to keep presentations of similar sentences apart, the sentences presented to a single participant were split up into two sequences that each contained only 1 old or new experimental sentence from each of the 12-sentence groups. For each participant, the recognition test started with a filler sentence. Then the two sequences of experimental and filler sentences were presented. The order of presentation again was pseudorandom with the restriction that no more than 2 filler sentences were presented successively.

Sentences were sampled read by a male speaker and played through a SoundBlaster AWE64 card. The auditory presentation and collection of responses was controlled by software using the PXL Psychological Experiments Library (Irtel, 1993).

Procedure. Participants were headphones during the study phase and the recognition test. In the study phase of the experiment, 18 sentences were presented 1 sentence at a time. Participants were given instructions similar to the instructions in the original experiment of Bransford et al. (1972): They were to listen carefully to the sentences, because they would be asked questions about them afterward. Each trial of the study phase started with the presentation of an exclamation point cue for 500 ms, cuing the auditory sentence presentation that followed. A sentence recording was played 800 ms after the onset of the cue. The participants were told to press a key as soon as they had understood the sentence. The cue for the next sentence was displayed 8 s after the response (in Bransford et al.’s experiment, a sentence was read to the participant every 10 s). Between cues, the screen remained blank. The presentation of a sentence took between 3.6 s and 5.1 s for experimental sentences (on average, 4.1 s and 4.0 s for different-models and same-model sentences, respectively) and between 2.4 s and 3.8 s for filler sentences.

Between the study phase and the recognition test, there was a pause of 3 min (as in the Bransford et al. ‘1972) experiment) in which the participants were instructed about the subsequent recognition test. In the recognition test, 30 sentences were presented. Participants were told to indicate as fast and as accurately as possible whether they had heard exactly the same sentence in the first part of the experiment. Each trial, as in the study phase, started with the presentation of an exclamation point cue for 500 ms. The sentence recording was played 800 ms after the onset of the cue. The participant answered with keys labeled yes and no and then gave a confidence rating on a 6-point scale ranging from very unsure to very sure. The given rating was displayed for 300 ms. Then the screen went blank for 3,200 ms until the cue for the next trial was given. The confidence ratings were explained during the instruction for the recognition test and were given by the number keys 1–6 on a standard keyboard labeled very unsure, unsure, tending to unsure, tending to sure, sure, and very sure.

Because it was possible that participants would be confused by the impression that sentences were repeated in the recognition test, the instruc-
tion for the recognition test, as in the Bransford et al. (1972) experiment, contained the information that some sentences may occur twice, but that participants should indicate by their decisions whether a sentence had occurred in the first part of the experiment. Actually, no sentence occurred twice. It took the participants approximately 15 min to complete the study phase and the recognition test, including the 3-min pause between the two.

Design. A full combination of the variables scene type (predator scene, peaceful scene), model type (same model, different models), and test probe type (old, new) yielded a $2 \times 2 \times 2$ factorial design. All varied within subjects, and each participant encountered three sentences of each combination in the recognition test.

Results

Recognition scores. As shown in Table 1, the predicted effect of causal relevance, reflected in a Scene Type $\times$ Model Type interaction, is confirmed by recognition scores. In Table 1, mean recognition scores for old and new experimental sentences are presented together with overall percentages of “old” responses (“yes” responses). Percentages of “old” responses are reported to provide information on recognition responses regardless of confidence ratings. Recognition scores were computed from old—new responses and confidence ratings. For old responses, they equal the confidence rating. For new responses, the confidence rating was multiplied by –1. Thus, the recognition scores vary between –6 for very sure new and 6 for very sure old. Positive recognition scores indicate correct recognition for old sentences. For new sentences, negative recognition scores reflect correct recognition judgments, and positive recognition scores reflect erroneous recognitions of new sentences as old.

The first two rows of Table 1 present recognition scores for peaceful scenes. For peaceful scenes, contrary to the model type effect in Bransford et al. (1972), new same-model sentences received higher negative mean recognition scores than did new different-models sentences and were less often erroneously recognized as old. In contrast, recognition scores for predator scenes indicate the model type effect as shown in Table 1. These recognition scores indicate more erroneous recognitions of new same-model sentences as old and better identification of new different-models sentences as new.

The Scene Type $\times$ Model Type interaction effect was confirmed in analyses of variance (ANOVA)s of recognition scores for new experimental sentences including the factors scene type and model type. $F_1(1, 31) = 6.65$, $MSE = 21.495$, $p < .05$, $f = .46$, with participants as a random factor, and $F_2(1, 22) = 5.85$, $MSE = 20.564$, $p < .05$, $f = .52$, with materials as a random factor nested within scene type. A significance level of .05 was chosen for all statistical tests, and $f$ denotes effect sizes following Cohen (1988). The main effects of scene type and model type were far from being statistically significant ($F < .02$).

In a second step, only recognition scores for new predator-scene sentences were focused in ANOVA.s. The model type effect was reflected in effect sizes analyzed according to participants, $F_1(1, 31) = 3.90$, $MSE = 20.010$, $p = .06$, $f = .36$, and analyzed according to materials, $F_2(1, 11) = 3.69$, $MSE = 24.210$, $p = .08$, $f = .58$, but did not reach statistical significance because of reduced power.

Discriminability scores. In discriminability scores, the model type effect is indicated by lower discriminability of same-model sentences compared with different-models sentences. As expected, discriminability scores presented in the top of Table 2 are consistent with a model type effect only for predator scenes. The discriminability score $d_e$ may be interpreted like $d'$ and captures recognition performance more completely than recognition scores for new sentences alone, because it is based on the error rates and confidence ratings for old and new sentences. For the computation of discriminability scores, the data from the group of 32 participants were treated as stemming from a single subject, because there were too few responses for a single-subject analysis (Macmillan & Kaplan, 1985). Frequencies of the 12 possible combinations of responses and confidence ratings ranging from very sure old to very sure new were computed for old sentences and new sentences.

For estimating $d_e$ I computed 11 different hit and false alarm rate ratios by assuming 11 hypothetical criteria, each of which divided the row of 12 frequencies into an old and a new part (Macmillan & Creelman, 1991). $RsCorePlus$ (Harvey, 2002) was used to estimate $d_e$ assuming normal distributions of noise, and noise with signal, and allowing for different standard deviations of the two distributions. $RsCorePlus$ uses a maximum-likelihood fitting algorithm and provides a standard error of the estimate of $d_e$. For some $d_e$ estimates, the goodness-of-fit was unsatisfactory. For these estimates, $d_e$ is reported as a rough indicator of overall recognition performance, no standard error ($SE_{d_e}$) is reported, and $d_e$ is set in parentheses in Tables 2 and 6. Discriminability scores were established separately for same-model, different-models, and filler sentences. As an alternative to $d_e$ that does not require distribution assumptions, the Goodman–Kruskal gamma coefficient was also calculated (Nelson, 1987). It yielded results similar to $d_e$ for all experiments.

Discriminability scores presented separately in Table 2 for peaceful- and predator-scene sentences clearly showed the expected scene type–model type interaction. With predator scenes, same-model sentences were less discriminable than were different-models sentences; the opposite was true of peaceful scenes, whose same-model sentences had a higher discriminability score than did different-models sentences. All discriminability scores of experimental sentences were low and reflected the difficulty of the recognition test. Given the standard errors for $d_e$ estimates within
Table 2

<table>
<thead>
<tr>
<th>Scene type</th>
<th>Sentence type</th>
<th>$d_a$</th>
<th>$SE_{d_a}$</th>
<th>$d_a$</th>
<th>$SE_{d_a}$</th>
<th>$d_a$</th>
<th>$SE_{d_a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peaceful</td>
<td>Same model</td>
<td>0.70</td>
<td>0.15</td>
<td>0.45</td>
<td>0.15</td>
<td>0.57</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Filler</td>
<td>(0.52)</td>
<td></td>
<td>0.59</td>
<td>0.15</td>
<td>(0.55)</td>
<td></td>
</tr>
<tr>
<td>Predator</td>
<td>Same model</td>
<td>0.61</td>
<td>0.16</td>
<td>0.69</td>
<td>0.16</td>
<td>0.65</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Filler</td>
<td>1.27</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>Same model</td>
<td>0.61</td>
<td>0.16</td>
<td>0.69</td>
<td>0.16</td>
<td>0.65</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Filler</td>
<td>1.27</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Parentheses indicate $d_a$ estimates with an unsatisfactory fit to the Gaussian model.

Scene type categories of 0.15, all satisfactorily estimated $d_a$ were significantly above zero. The difference between $d_a$ scores had to be at least 0.42 to be statistically significant (e.g., Masson & Loftus, 2003).

Peaceful-scene same-model sentences had the highest discriminability score among the experimental sentences (0.70). This was unexpected and may have been due to the differences in interpretations of spatial expressions between peaceful and predator scenes, which I consider in the General Discussion. The discriminability score of the filler sentences, which had more discrepancies within a pair, was higher than the discriminability score for experimental sentences, as expected.

Study times and response times in the recognition test. Mean study times for experimental and filler sentences measured 8.0 s and 6.8 s from the start of auditory presentation, respectively ($SDs = 2.7$ and 2.4). There were no differences in mean study times, even for combinations of model type and scene type, that would yield an explanation of the Model Type × Scene Type interaction in terms of study duration.

In the recognition test, the mean response times for experimental and filler sentences were 7.3 s and 6.1 s, respectively ($SDs = 1.2$ and 1.4). Again, there were no hints at alternative explanations of the Model Type × Scene Type interaction in recognition performance.

Discussion

Recognition scores and discriminability scores confirmed the predicted Model Type × Scene Type interaction. Causal relevance of space was signaled only by predator-scene sentences. For predator scenes only, discriminability of different-models sentences was better than discriminability of same-model sentences. Compared with Bransford et al.’s (1972) experiment, the model type effect found with predator scenes is less clear. In Bransford et al.’s experiment, the recognition scores indicated participants’ complete failure to discriminate between old and new same-model sentences. Nonetheless, because Experiment 1 was designed as a close replication of the experiment by Bransford et al. with systematic variation of scene type, the strong Scene Type × Model Type interaction supports the presumption that causal relevance of space might have been crucial for the original finding. Presumably, most of the original sentence material resembled the predator-scene material in the present study and signaled causal relevance of space. If this was actually the case, the classic evidence for spatial situation model construction, without explicit instruction, would fit better with present theories that rate causal relevance as an important trigger for spatial situation model construction during text comprehension (e.g., Radvansky & Copeland, 2000).

In Experiment 1, causal relevance of space was signaled by predator scenes and had to be detected by participants, similar to naturalistic text comprehension. In Experiments 2–4, a danger judgment task was used in the study phase, which was supposed to direct participants to evaluate causal relevance of space even if it was not signaled by the text. The evaluation of causal relevance induced by the danger judgment task was supposed to indirectly trigger the construction of spatial situation models even for peaceful scenes. Consequently, the model type effect was expected for predator and for peaceful scenes alike in Experiments 2–4.

In addition to study condition, presentation modality and ease of encoding were varied in Experiments 2–4. In Experiment 2, an auditory presentation like the one in Experiment 1 was used together with the danger judgment task in the study phase. However, double auditory presentation was used in Experiment 2, because a single auditory presentation turned out to be too elusive when the danger judgment covered the recognition test.

Experiment 2: Danger Judgment and Double Auditory Presentation

Experiment 2 was similar to Experiment 1 except for the study phase. Participants were not simply instructed to listen to the sentences in the study phase but were asked to judge whether an animal in the described scene in each sentence was in danger. This danger judgment task was introduced for two purposes. First, it allowed for a more direct test of the supposed link between causal relevance of space and the spontaneous construction of spatial situation models. If I could show that comprehenders, directed to evaluate causal relevance, are more likely to construct mental models of causally relevant spatial relations, I would have evidence for the effect of causal relevance. With the danger judgment task, a model type effect was also expected for peaceful scenes, because participants had to expect causal relevance of space even for peaceful scenes, although spatial relations were never causally relevant in peaceful scenes. The expectation was created by the filler sentences they encountered. In some filler sentences, danger did not result from predator animals but from the described spatial relations, for example, A donkey steps on a hill in which a mole digs (F5, see the Appendix). Therefore, participants could not infer
absence of danger from the absence of a predator animal. They had to consider spatial relations even for peaceful-scene sentences.

The second purpose I pursued by using the danger judgment task was covering the recognition test. The danger judgment task ensured that participants processed the study sentences for comprehension, but it made it unlikely that the sentences were memorized intentionally. Without a cover task, as in Experiment 1 and Bransford et al.'s (1972) experiment, participants may intentionally construct visuospatial representations as a strategy for memorization prompted by the predominance of spatial descriptions in the sentence material. However, with the danger judgment task as cover task, a model type effect can be more confidently attributed to causal relevance.

The danger judgment task effectively covers the recognition test. This was shown in a pilot experiment with single auditory presentation and the danger judgment task, which is not reported in detail. In this experiment, recognition performance was impaired and provided only slight evidence for a model type effect ($d_p$ was 0.14 vs. 0.19 for same model and different models, respectively). A single auditory presentation combined with the danger judgment task yielded representations that were useless for discriminating between the very similar experimental sentences of a same-model or different-models pair. The danger judgment task covered the recognition test, and there was no reason for participants to intentionally memorize the sentences. To alleviate the drop in recognition performance, which presumably prevented a clearer model type effect, I ensured that recognition was supported by a repetition of auditory presentation in the study phase of Experiment 2. This was expected to counteract the elusive character of auditory presentation and should provide the prerequisite for a model type effect.

**Method**

**Participants.** Thirty-two students from the University of Regensburg and of the Chemnitz University of Technology, Chemnitz, Germany, 10 of whom were psychology students, participated in the experiment either to fulfill a curricular requirement or for monetary compensation that was equivalent to $5.

**Materials.** The same sentences that were used in Experiment 1 were used in Experiment 2.

**Procedure.** The procedure resembled the procedure in Experiment 1, except that each sentence presentation in the study phase was immediately followed by a second presentation of the same sentence, and participants gave a danger judgment after the double auditory presentation. Participants were instructed to judge each sentence according to whether an animal or a group of animals was in immediate danger in the described scene. Two sentences describing scenes with danger presence and danger absence that were not used in the experiment were given as examples in the instruction. Half of the experimental sentences described predator scenes with predator animals that could harm other animals, and half described peaceful scenes with animals that could not harm one another.

How dangerous predator animals were to other animals in the predator scenes depended on spatial relations. Thus, participants had to consider spatial relations to judge danger. Danger judgments are open to interpretation. For example, the zebras in *Two zebras move away from a shrub and a lion trots towards it* may be seen as facing danger, or they may be seen as safe, because the lion seems to ignore them. Therefore, the danger present in experimental scene descriptions could only be balanced based on plausible precategorizations. I precategorized the sentences with two colleagues. Filler sentences that also varied with regard to danger presence in

the described scenes were chosen to approach an even distribution of expected danger judgments and no danger judgments for each participant. For all the participants, the first sentence in the study phase was a filler sentence in which a peaceful animal caused danger for a predator animal. The other filler sentences also described scenes in which peaceful animals could harm one another or could even harm predator animals because of spatial relations. This should have discouraged participants from simply relying on the presence of a predator animal when judging the threat to animals in the experimental sentences.

Each trial of the study phase started with the presentation of an exclamation point on the computer screen for 500 ms cuing the subsequent auditory sentence presentation. A sentence recording was played twice 600 ms after onset of the cue. The interval between the two presentations was 1,000 ms, and the complete double presentation lasted 8.2–11.1 s for experimental sentences (on average, 9.2 s and 9.0 s for different-models and same-model sentences, respectively) and 5.8–8.5 s for filler sentences. In the recognition test, as in Experiment 1, only single auditory presentations preceded the recognition responses and confidence ratings.

During and after the auditory presentation, the screen remained blank until the participant responded with a judgment of danger in the described scene. Participants were instructed to listen to both presentations before responding and to answer as quickly and as accurately as possible. The response, *yes* or *no*, was displayed and remained visible for 300 ms. Then the screen went blank for 7,200 ms, followed by the presentation of an exclamation point signaling the next sentence presentation. After the study phase, there was a pause of 3 min followed by the recognition test. The recognition test proceeded as in Experiment 1.

**Results**

**Judgments of danger.** In the study phase, the participants judged the danger present in the scenes described by the peaceful, predator, and filler sentences. To confirm that participants had not judged danger based simply on the presence or absence of a predator animal, participants' judgments were compared with the precategorizations.

Of all the filler sentences precategorized as dangerous, 80% were judged accordingly by participants. The agreement with the precategorizations of predator-scene sentences as dangerous was 82%, while agreement with the precategorizations of predator-scene sentences as not dangerous was 39%. This indicates that the presence of a predator animal in a described scene strongly drew participants to judging the scene as dangerous; nevertheless, other considerations, presumably including spatial relations, clearly influenced participants' danger judgments. Therefore, the intended prerequisite for a model type effect on recognition performance for both scene types has been established. Similar rates for danger judgments were obtained in the following experiments. To avoid redundancy, results of danger judgments are not reported for the following experiments.

**Recognition scores.** The double auditory presentation, as expected, saved recognition performance and established the prerequisite for a model type effect. Recognition performance with the danger judgment in the study phase that covered the recognition test improved compared with the briefly mentioned pilot study with single auditory presentation. A clear model type effect was present in mean recognition scores, as is shown in Table 3. Unexpectedly, the model type effect was mainly due to the recognition of new peaceful-scene sentences. For new peaceful-scene
sentences, the mean recognition score for same-model sentences was considerably higher than it was for new different-models sentences (2.13 vs. –0.05, respectively). For the predator-scene sentences, the mean recognition score for new same-model sentences was only slightly higher than it was for new different-models sentences (0.05 vs. –0.14, respectively). The model type main effect was, nonetheless, statistically significant in ANOVAs including the variables model type and scene type, \(F(1, 31) = 6.05, \text{MSE} = 18.232, p < .05, f = .44,\) and \(F(1, 22) = 5.43, \text{MSE} = 22.590, p < .05, f = .50.\) The Model Type \(\times\) Scene Type interaction was not statistically significant analyzed according to participants, \(F(1, 31) = 3.04, p = .09, f = .31,\) but the main effect of scene type was, \(F(1, 31) = 7.24, p < .05, f = .48.\) The analyses according to materials yielded \(F(1, 22) = 4.37, p < .05, f = .45\) and \(F(1, 22) = 1.89, p = .18, f = .29,\) respectively. As discussed below, the danger judgment task seems to have oriented participants differently toward predator- and peaceful-scene sentences. Nonetheless, compared with the following experiments, it is exceptional that the model type effect was nearly absent for predator scenes.

**Discriminability scores.** Discriminability scores shown in Table 2 confirmed the clear model type effect for peaceful scenes. Same-model discriminability for peaceful scenes was nearly zero (0.03); different-models discriminability was reliably higher (0.61); the critical difference is 0.45 given the standard errors of 0.16. For predator scenes, discriminability scores were the same for both model types. The total \(d_{z}\) scores calculated for recognition responses pooled over scene types differed by 0.30 (the critical difference corresponding to the standard error of 0.11 is 0.31). The discriminability of filler sentences was similar to the filler discriminability in Experiment 1.

**Response times.** Mean response times in the study phase were measured from the start of the first auditory presentation and were only 0.7 s longer than the complete duration of the double auditory presentation for experimental sentences. Obviously, participants often had made their decision before the end of the second presentation. The mean response time in the study phase was 9.8 s for experimental sentences and 7.9 s for filler sentences (SDs = 0.3 and 0.4). In the recognition test, mean response times measured from the start of a single auditory presentation were 5.9 s for experimental sentences and 4.8 s for filler sentences (SDs = 0.9 and 0.7). Again, there were no hints to explanations of the recognition response pattern in terms of response times.

**Discussion**

The results confirm the predicted effect of double auditory presentation in the study phase. Overall recognition performance improved compared with a pilot experiment with single auditory presentation and the danger judgment task. Double auditory presentation seems to be a critical precondition for a clear model type effect, with the danger judgment task covering the recognition test. Supported by two auditory presentations, recognition performance was high enough to display effects of model type.

The clear model type effect for peaceful scenes is in accordance with the prediction that the danger judgment task would induce the construction of spatial situation models also for peaceful scenes. But, unexpectedly, I found no model type effect for predator scenes, although causal relevance of space was signaled by predator scenes, as it was in Experiment 1. The absence of a model type effect for predator scenes in this experiment is exceptional. In Experiments 1, 3, and 4, a model type effect was found for predator-scene recognition scores, even in the danger judgment condition of Experiment 3, which is an exact replication of Experiment 2. However, the pattern of a stronger model type effect for peaceful scenes than for predator scenes and nearly absent discriminability of old and new peaceful scene same-model sentences was reoccurring in experiments that also employed the danger judgment task and provided support for extended encoding. The danger judgment task seems to orient participants differently toward the two scene types.

Predator-scene sentences signal possible danger and ensure their thorough evaluation. Possibly, besides spatial relations, participants evaluate inferences for predator-scene sentences by drawing on background knowledge. For example, participants could think about whether the peaceful animals are natural prey for the respective predator animals, or they could try to infer probable intentions of predator animals from the described movements. If this richer encoding enhances memory for surface features of the sentences, it improves recognition. Additionally, the interpretation of spatial relations could have differed for the two scene types as explained in the General Discussion.

In contrast, for peaceful scenes, there is less to evaluate besides spatial relations. A focus on the spatial mental model yields no memory for surface features. Consequently, responses to old and new same-model sentences are similar in the recognition test. To clarify the effects of study conditions, in Experiment 3, a mere study condition was contrasted with a danger judgment study condition.

**Experiment 3: Mere Study Versus Danger Judgment, Double Auditory Presentation**

Experiment 1 and Bransford et al.’s (1972) study used single auditory presentation. With the danger judgment task covering the recognition test, double auditory presentation was necessary in Experiment 2. Therefore, double auditory presentation was chosen for the variation of study condition in a single experiment. If
double auditory presentation does not change the presumed effect of scene type in a mere study condition, the model type effect should only occur for predator scenes in the mere study condition. But it should be present for predator and peaceful scenes in the danger judgment study condition. These predictions would be confirmed by a corresponding Study Condition × Scene Type × Model Type three-way interaction.

Method

Participants. Sixty-four students of the Chemnitz University of Technology, 26 of whom were psychology students, participated in the experiment either to fulfill a curricular requirement or for monetary compensation (equivalent to $5). Half of the participants were randomly assigned to the mere study condition, and the other half gave danger judgments in the study phase.

Materials. The same sentences that were used in Experiments 1 and 2 were used in Experiment 3.

Procedure. The procedure in the danger judgment condition resembled the procedure in Experiment 2. In the mere study condition, the procedure was similar to the procedure in Experiment 1, except that this time each sentence was presented twice in the study phase. The interval between the auditory presentations in the study phase was 1,000 ms in both conditions. The recognition test in both study conditions was the same as in Experiments 1 and 2.

Design. The variables scene type (predator scene, peaceful scene), model type (same model, different models), and test probe type (old, new) were fully combined and varied within subjects, and study condition (mere study, danger judgment) was a between-subjects variable. This yielded a $2 \times 2 \times 2$ mixed factorial design.

Results

A three-way Study Condition × Scene Type × Model Type interaction was predicted, because the model type effect should be present for all the combinations of study condition and scene type, except for peaceful scenes in the mere study condition. The predicted three-way interaction was not confirmed by analyses of recognition scores for new sentences alone. But it was shown by discriminability scores, which include responses to old sentences, and by an index of confident correct responses to old and new sentences.

Recognition scores. Mean recognition scores for new sentences, as shown in Table 4, were consistently higher for same-model sentences than they were for different-models sentences; thus they indicated a general model type effect. Overall percentages of “old” responses concur, except for peaceful scenes in the danger judgment condition. The model type main effect was noticeable, yet not statistically significant in ANOVAs of new scene-identification scores including the variables model type, scene type, and study condition, $F_1(1, 62) = 3.91, MSE = 19.106, p = .05, f = .25$, with participants as a random factor nested within study condition, and $F_2(1, 22) = 3.57, MSE = 21.582, p = .07, f = .40$, with materials as a random factor nested within scene type. In the danger judgment study condition, the peaceful-scene recognition scores for new sentences were higher than for the remaining combinations of study condition and scene type, which corresponds to a Scene Type × Study Condition interaction that was statistically significant analyzed according to participants, $F_1(1, 62) = 4.13, p < .05, f = .26$, and $F_2(1, 22) = 1.05, p = .32, f = .22$. The $2 \times 2 \times 2$ ANOVAs yielded no other significant main effects or interactions.

Pooled over scene types and analyzed separately for study conditions, the model type effect on recognition scores of new experimental sentences was smaller for the danger judgment study condition than it was for the mere study condition. In ANOVAs for the danger judgment study condition, the model type effect did not reach statistical significance, $F_1(1, 31) = 3.13, MSE = 13.634, p = .09, f = .32$, and $F_2(1, 23) = 1.15, MSE = 37.237, p = .30, f = .22$. For the mere study condition, the model type effect was confirmed by analyses according to participants and according to materials, $F_1(1, 31) = 4.50, MSE = 22.013, p < .05, f = .38$, and $F_2(1, 23) = 4.78, MSE = 20.730, p < .05, f = .46$.

Discriminability scores. The predicted three-way interaction was shown by discriminability scores, which are presented in Table 2. They indicated a model type effect for peaceful and predator scenes in the danger judgment study condition, and for predator scenes in the mere study condition but not for peaceful scenes in the mere study condition. For the danger judgment study condition, which replicated Experiment 2, discriminability scores showed a model type effect for predator scenes, too. As in Experiment 2, peaceful-scene discriminability scores were lower than predator-scene discriminability scores. Except for peaceful-scene same-model discriminability in the danger judgment condition, all satisfactorily estimated discriminability scores were reliably different from zero. The difference between same-model and different-models discriminability scores was smaller than the critical difference corresponding to the respective standard errors; therefore discriminability scores provided no sufficient statistical evidence for the predicted three-way interaction. As an alternative, correct confident responses were analyzed.
Correct confident responses. The predicted three-way interaction was displayed by the rates of those correct recognition responses that were assigned the highest confidence rating by participants. To arrive at an index, those very sure correct responses to old and new sentences were assigned 1, and all other responses were assigned 0. Means were calculated for each participant and then multiplied by 100. The index thus corresponds to the percentage of all responses to old and new sentences that were correct and of which participants were very sure. The means of these percentages were higher for different-models sentences than they were for same-model sentences in the danger judgment study condition, indicating the model type effect (24.2 vs. 11.7 for peaceful scenes, 25.8 vs. 19.9 for predator scenes, SEs = 3.9, 3.9, 4.1, and 4.3, respectively). Different-models mean percentages were also higher for predator scenes in the mere study condition (30.9 vs. 25.4, SEs = 4.8 and 4.0) but lower than same-model mean percentages for peaceful scenes (24.2 vs. 30.1, SEs = 3.8 and 4.4).

The corresponding predicted three-way interaction was close to statistical significance in an ANOVA with scene type and model type as within-subjects variables and study condition as between-subjects variable, F(1, 62) = 3.92, MSE = 329.31, p = .05, f = .25. The Study Condition × Model Type interaction was significant, F(1, 62) = 4.68, MSE = 300.49, p < .05, f = .27, as was the main effect of model type, F(1, 62) = 4.30, p < .05, f = .26.

Response times. Mean study times in the danger judgment study condition were around 2.3 s shorter than they were in the mere study condition (9.7 s vs. 12.2 s for experimental sentences, 7.9 s vs. 10.0 s for filler sentences, SDs = 0.4, 2.7, 0.5, and 2.5). Mean response times in the recognition test were only around 0.9 s shorter in the danger judgment condition (6.4 s vs. 7.2 s for experimental and 5.5 s vs. 6.5 s for filler sentences, SDs = 1.1, 1.0, 1.1, and 1.6).

Discussion

Discriminability scores showed the predicted three-way interaction. It was close to statistical significance, if the correct responses that were focused on were those of which participants were very sure. However, analyses of recognition scores for new sentences alone pointed to a general model type effect.

For the danger judgment study condition, discriminability scores were in accordance with results of new sentence-recognition score analyses. Both provided some evidence for a model type effect that affected recognition performance for peaceful scenes and for predator scenes, although the evidence was weak in terms of statistical significance. In contrast to Experiment 2, which the danger judgment study condition exactly replicated, this time the model type effect occurred for both scene types as predicted. Similar to Experiment 2, the overall discriminability of peaceful-scene sentences was lower than it was for predator-scene sentences. Again, the danger judgment task probably oriented participants differently toward peaceful scenes and predator scenes.

For the mere study condition, analyses of recognition scores for new sentences alone suggest a model type effect for both scene types. However, in discriminability scores and percentages of correct confident responses, the model type effect was limited to predator scenes, as in Experiment 1. Discriminability scores include responses to old sentences and therefore provide a complete picture of response behavior. Percentages of correct confident responses also include responses to old sentences. Nonetheless, the results would have been more conclusive if recognition scores for new sentences alone had shown the predicted scene type difference. In summary, it can be stated that the model type effect was larger for predator scenes than it was for peaceful scenes in the mere study condition.

In Experiment 1’s mere study condition and single auditory presentation resembling the Bransford et al. (1972) experiment, the model type effect was absent for peaceful scenes. Double auditory presentation could have induced a weak model type effect for peaceful scenes in the mere study condition of Experiment 3 in contrast to single auditory presentation in Experiment 1. The double auditory presentation in Experiment 3 could have enhanced the probability of model construction for peaceful scenes by either easing model construction or providing more time for a deliberate decision to construct a model, or both.

So far, presentation of sentences had been auditory as in the Bransford et al. (1972) experiment. In Experiment 4, the presentation modality was changed. Visual presentation was used to test predictions concerning effects of presentation modality and ease of encoding on the probability of spontaneous spatial situation model construction. The interference hypothesis states that spatial representation is more probable with auditory presentation, because with visual presentation, the reading and construction of spatial representations would compete and interfere. In the Bransford et al. study, presentation was auditory. Presentation modality, therefore, was suggested as one explanation for the reduced effects in a replication of Bransford et al. (1972) experiment, which used visual presentation and seems to be the first published replication (Rinck et al., 2001).

Experiment 4: Self-Paced Visual Presentation With Danger Judgment

Experiment 4 was designed to resemble Experiments 1–3, except for the presentation modality, which was now visual. The danger judgment task was included in the study phase; thus a model type effect was expected again for predator and peaceful scenes. With the change to self-paced visual presentation, I explored the effect of presentation modality on the probability of spontaneously constructed spatial representations. Bransford et al.’s (1972) study is remarkable not only because the participants obviously constructed spatial representations without having been asked to but also because presentation was auditory. The majority of current text-comprehension studies use visual presentation.

Modality of Presentation and Ease of Encoding

Situation-model theories of text comprehension usually do not specify how spatial relations are represented in situation models (Zwaan & Radvansky, 1998). But predominantly spatial situation models are often described as analogue visuospatial representations (e.g., Fincher-Kiefer, 2001; Knauff & Johnson-Laird, 2002; Langston et al., 1998). Brain imaging studies (Mellet et al., 1996) and experiments on embodied cognition (e.g., Glenberg & McDani, 1992; Stanfield & Zwaan, 2001) support this assumption. The plain-interference hypothesis states that spontaneous construction of spatial representations should be easier and more probable with auditory presentation, because reading is known to potentially
compete and interfere with the construction of visuospatial representations (Brooks, 1970; Glass, Millen, Beck, & Eddy, 1985; Hörnig, Claus, & Eyferth, 2000; Kaup, Kelter, & Habel, 1999). As mentioned in Experiment 3’s Discussion section, Rinck et al. (2001) presume that auditory presentation might have been one of the reasons for the reduced model type effect in their replication, which used visual presentation, different sentence material, and a memory instruction. If this assumption were correct, a change to visual presentation would reduce the model type effect on recognition performance. However, at least two qualifications of the assumed effect of presentation modality seem necessary.

First, an interference of a visual presentation of text with the construction of spatial situation models presumably occurs only if there is time pressure during reading. If readers have enough time to pause their reading in order to construct spatial situation models, they probably can avoid interference effects. Second, interference might not only be easily avoided, but visual presentation may even support the spatial situation model construction by providing an external representation of text. It may be easier to construct spatial representations from a written sentence if the sentence can be read repeatedly, thereby making a stepwise construction of spatial representations possible. It is known that repeated encoding of spatial descriptions fosters the construction of spatial representations from text (e.g., Baguley & Payne, 2000; Denis & Cocude, 1992; Perrig & Kintsch, 1985). Of course, this advantage of visual presentation will only be effective if comprehenders have time and decide to reread.

**Method**

**Participants.** Thirty-two students from the University of Regensburg, 20 of whom were psychology students, participated in the experiment either to fulfill a curricular requirement or for monetary compensation that was equivalent to $5.

**Materials.** The same sentences that were used in Experiments 1–3 were used in Experiment 4.

**Procedure.** The procedure was similar to the procedures used in Experiments 1–3, except that the sentences were presented visually. They were shown in white on a black computer screen, again controlled by software using the PXL Psychological Experiments Library (Irlt, 1993). Presentations in the study phase and the recognition test were self-paced.

The study phase included the danger judgment task. Each trial of the study phase started with the presentation of a sentence that was displayed until the participant responded. A display of the “yes” or “no” response replaced the sentence and remained visible for 300 ms. Then the screen went blank for 8 s, followed by the presentation of the next sentence.

After the study phase, there was a pause of 3 min. For the recognition test, participants were instructed to indicate as fast and as accurately as possible whether they had read exactly the same sentence in the first part of the experiment. As in Experiments 1–3, they answered with keys labeled yes and no, and a confidence rating. In the recognition test, each sentence remained visible until the participant responded yes for old and no for new. After the response, the sentence was replaced by How sure? and the participant gave a confidence rating. The rating was displayed for 300 ms. Then the screen went blank, and the next trial started 8 s later with the presentation of the next sentence.

**Design.** The variables scene type (predator scene, peaceful scene), model type (same model, different models), and test probe type (old, new) yielded a $2 \times 2 \times 2$ factorial design. They were all varied within subjects.

**Results**

**Recognition scores.** Mean recognition scores for new experimental sentences presented in Table 5 indicated the model type effect. It was stronger than it was with double auditory presentation and stronger than the model type effect found with a single auditory presentation in the pilot study. ANOVAs of recognition scores for new sentences including the factors scene type and model type confirmed the main effect of model type, $F_1(1, 31) = 11.97, MSE = 19.85, p < .01, f = .62$, and $F_2(1, 22) = 14.14, MSE = 19.588, p < .01, f = .80$. There was no statistically significant evidence for a Scene Type × Model Type interaction ($F_1$ and $F_2 < 0.05) or for a main effect of scene type, $F_1(1, 31) = 2.70, p = .11, f = .29; F_2(1, 22) = 0.29, p = .59, f = .11$. Only small differences in recognition scores were related to scene type. The model type effect was stronger in mean recognition scores for new peaceful-scene sentences (0.97 vs. –0.73 for same-model vs. different-models sentences, respectively) than it was in mean recognition scores for new predator-scene sentences (0.16 vs. –1.38). This scene type difference was similar to the difference in Experiments 2 and 3 with the danger judgment and double auditory presentation.

**Discriminability scores.** Discriminability scores presented in Table 6 further confirmed the model type effect. The model type effect displayed by total discriminability scores was similar to that obtained with double auditory presentation, but the difference between total scores for same-model and different-models sentences of 0.26 was below the critical difference (0.31 for SEs of 0.11).

**Response time.** Mean response times for the danger judgments in the study phase measured from the beginning of presentation were 6.4 s for experimental sentences and 4.7 s for filler sentences (SDs = 2.7 and 2.0). Mean response times in the recognition test were 8.0 s for experimental sentences and 6.1 s for filler sentences (SDs = 2.3 and 1.8).

**Discussion**

The predictions based on the expected spontaneous construction of causally relevant spatial situation models were again confirmed.

<table>
<thead>
<tr>
<th>Scene type</th>
<th>Old M</th>
<th>Old SE</th>
<th>Old %</th>
<th>New M</th>
<th>New SE</th>
<th>New %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peaceful scenes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same model</td>
<td>1.91</td>
<td>0.45</td>
<td>64.6</td>
<td>0.97</td>
<td>0.49</td>
<td>62.5</td>
</tr>
<tr>
<td>Different models</td>
<td>1.68</td>
<td>0.57</td>
<td>67.7</td>
<td>–0.73</td>
<td>0.52</td>
<td>39.6</td>
</tr>
<tr>
<td>Predator scenes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same model</td>
<td>2.45</td>
<td>0.41</td>
<td>70.8</td>
<td>0.16</td>
<td>0.64</td>
<td>53.1</td>
</tr>
<tr>
<td>Different models</td>
<td>1.71</td>
<td>0.48</td>
<td>63.5</td>
<td>–1.38</td>
<td>0.52</td>
<td>40.6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same model</td>
<td>2.07</td>
<td>0.36</td>
<td>67.7</td>
<td>0.84</td>
<td>0.42</td>
<td>57.8</td>
</tr>
<tr>
<td>Different models</td>
<td>1.69</td>
<td>0.44</td>
<td>65.6</td>
<td>–0.95</td>
<td>0.44</td>
<td>40.1</td>
</tr>
</tbody>
</table>
The danger judgment task was effective in inducing the consideration of the potential causal relevance of spatial situations described by peaceful-scene and predator-scene sentences. It indirectly triggered the construction of spatial situation models that helped to discriminate different-model sentences but not same-model sentences in the surprise recognition test. The model type effect was confirmed by recognition scores and discriminability scores and was stronger than in the experiments with double auditory presentation.

With respect to the interference hypothesis versus the prediction of eased encoding and eased model construction, the results provided no decisive evidence. However, the outcome is more problematic for the plain-interference hypothesis, which would have predicted a reduced model type effect compared with the experiments with auditory presentation.

Defending the interference hypothesis, the stronger model type effect compared with auditory presentation may be explained by a simple support of sentence recognition. A simple support of sentence recognition by visual presentation may have been caused by differences between typographic images of old and new sentences from a sentence pair, which may have helped in discriminating between very similar sentences (Flagg & Reynolds, 1977). It is possible that the small model type effect with a single auditory presentation and the danger judgment in the pilot experiment resulted from low overall recognition performance. Thus, the strong model type effect in Experiment 4 with visual presentation may only indicate eased encoding, not easier spatial situation model construction with visual presentation.

Notwithstanding this alternative account of the strengthened model type effect, the results showed effects of spatial situation models on recognition performance under conditions of visual presentation. The construction of spatial situation models that was required for the danger judgment might compete and interfere with reading, but, presumably, noticeable interference effects occur only under time pressure. The results render it unlikely that interference explains difficulties replicating the Bransford et al. (1972) experiment with extended visual presentation (20 s in Rinck et al., 2001). In fact, in a further unreported experiment that was similar to Experiment 4 except for restricted visual presentation, I found an only slightly smaller model type effect ($d_a$ = 0.63 vs. 0.43) with visual presentation times of 4.2–7.2 s for experimental sentences (300 ms per syllable). These presentation times were chosen to allow for one slow reading pass.

The results with visual presentation, together with the results of Experiments 1–3, point to the importance of eased encoding for spatial situation model construction. Causal relevance is probably more important than presentation modality in finding effects of spontaneous model construction in the recognition paradigm of Bransford et al. (1972).

General Discussion

My main objective in this study was to demonstrate the critical role of causal relevance in increasing the probability that comprehenders would spontaneously construct spatial situation models. The classic finding of Bransford et al. (1972) was replicated. Different-models sentences were easier to discriminate in a recognition test than same-model sentences. However, this model type effect was less clear than it was in the Bransford et al. experiment. In Experiment 1, with a mere study instruction and with a single auditory presentation, the model type effect was found only for sentence material that signaled causal relevance of space (predator scenes). In Experiment 3, with a mere study condition and with double auditory presentation, it was stronger for predator scenes. In Experiments 2–4, with the danger judgment task in the study phase, the evaluation of causal relevance led to the construction of spatial situation models and consequently produced the model type effect, even for material that did not signal causal relevance. Experiment 1 was a close replication of Bransford et al.'s experiment and, together with the results of Experiments 2–4, suggests a revised interpretation of the classic study.

First, causal relevance, which is clearly present in the less prominent sample material from the Bransford et al. (1972) study, seems to have been crucial for the original finding. Given this, the original finding seems less exceptional compared with later studies on spatial situation models that did not yield evidence for spontaneous construction without special tasks or instructions. It now fits nicely with several recent demonstrations of the important role of causal relevance plays in spontaneous spatial representation in text comprehension (Levine & Klin, 2001; Radvansky & Copeland, 2000; Radvansky et al., 2003; Rice, 1999; Sundermeier et al., 2004).

Second, foregrounding of space is not a sufficient explanation of spontaneous construction in the Bransford et al. (1972) study, as it would predict a model type effect for peaceful scenes in Experiment 1. Also, in several other studies, foregrounding of space did not suffice to induce spontaneous spatial representation (de Vega, 1995; Hakala, 1999; Levine & Klin, 2001; Zwaan & van Oostendorp, 1994).

And finally, the modality of presentation cannot solely predict construction activity. The model type effect on recognition performance, which indicates spatial situation model construction, was similar with double auditory presentation and self-paced visual presentation (Experiments 2–4). Therefore, it was not just modality but the opportunity for extended encoding provided with visual or auditory presentation that was crucial. The model type effect was nearly absent, and recognition performance broke down with a single auditory presentation in a pilot experiment. The systematic variation of presentation modality and ease of encoding demonstrated that, contrary to the plain-interference hypothesis, visual presentation can strengthen the model type effect if encoding is easy. The plain-interference hypothesis has to be qualified by ease

Table 6

Discriminability Scores $d_a$ With Standard Errors (Experiment 4, Self-Paced Visual Presentation, Danger Judgment, $N = 32$)

<table>
<thead>
<tr>
<th>Scene type</th>
<th>Sentence type</th>
<th>Peaceful</th>
<th>Predator</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d_a$</td>
<td>$SE_{da}$</td>
<td>$d_a$</td>
<td>$SE_{da}$</td>
</tr>
<tr>
<td>Same model</td>
<td>0.18</td>
<td>0.15</td>
<td>0.44</td>
<td>0.15</td>
</tr>
<tr>
<td>Different models</td>
<td>0.51</td>
<td>0.15</td>
<td>0.64</td>
<td>0.15</td>
</tr>
<tr>
<td>Filler</td>
<td>(1.21)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Parentheses indicate a $d_a$ estimate with an unsatisfactory fit to the Gaussian model.
of encoding. Interference of reading and construction seems to occur only under time pressure, at least for simple spatial situation models. Consequently, visual presentation is not a sufficient explanation for the difficulty of replicating the Bransford et al. (1972) study and is not a likely explanation for the difficulties of finding spontaneous construction of spatial situation models.

For the spontaneous construction of spatial situation models, causal relevance has to be considered together with other important preconditions, as listed in the introduction to this article. All of these important preconditions concerning ease of construction and comprehender’s expectation were fulfilled in the present experiments and in the Bransford et al. (1972) study. The experimental sentences located only three entities and followed conventions of spatial descriptions. The beneficial effect of extending encoding opportunities was shown with self-paced visual presentation in Experiment 4 and with double auditory presentation in Experiments 2 and 3. Space was foregrounded in the materials, but this alone did not suffice to produce the model type effect. Causal relevance of space was necessary, either signaled by predator scenes in Experiment 1 or stressed by the danger judgment task in Experiments 2–4. And finally, sentences, as in the predator-scene examples from the original study, remained indeterminate with regard to potential interactions of animals in the scenes and motivated participants to infer probable continuations.

These preconditions for the spontaneous construction of spatial situation models in the reported experiments can be summarized as simple, conventionally described configurations of few entities, causal relevance, and modest “suspense”. Similar preconditions were seldom fulfilled by materials and tasks in experimental studies. The evidence for spontaneous construction under these preconditions is consistent with constructionist theories of text comprehension (Gaesser & Kreuz, 1993) and also with the minimalist position (McKoon & Ratcliff, 1992), which states that inferences are only drawn if they are easy and central to comprehension. It was confirmed that comprehenders invest effort strategically.

However, it is not justified to conclude from many failures to demonstrate spontaneous construction in experiments that comprehenders normally refrain from representing space during naturalistic text comprehension. The preconditions for spontaneous construction, including causal relevance, are not uncommon outside the laboratory and yield the common experience of visuospatial representations during comprehension of narrative texts. Causal relevance is an important precondition and may even be crucial. Additionally, the study of spontaneous construction requires that participants’ spatial representations are not directly influenced and unified by tasks and instructions.

The better discrimination of different-models sentences that constitutes the model type effect may have resulted from differing spatial situation models and from differing traces of the model construction process. Situation models presumably were constructed in the study phase and in the recognition test. From study to test, the situation model constructed in the study phase had to be retained in long-term memory. The long-term memory representation of spatial mental models may take different forms. Baguley and Payne (2000) have shown that spatial mental models may be retained as a construction trace, but also, or additionally, as an image. Therefore, the component of the recognition judgment at test that was responsible for the model type effect may have been based on the familiarity of the construction process, the familiarity of the constructed visuospatial representation, or both. Recognition judgments and confidence ratings were also, presumably, affected by more general phenomena of recognition memory. Participants could use episodic content in deciding whether they had encountered a sentence in the study phase. Episodic content from the study phase that fitted participants’ experience of new sentences in the recognition test may explain the often high confidence in erroneous recognition of new sentences as old, which is typical of false memories. The high false memory rates in the present experiments are not uncommon for difficult sentence-recognition tasks (Holmes, Waters, & Rajaram, 1998).

Scene Type Effects

Scene type, apart from the expected interaction with model type in Experiment 1 and the less clearly confirmed interaction in Experiment 3, also had unexpected effects. The suggested explanations are post hoc and in part rather subtle, but may help in conducting similar studies. The scene type and danger variation itself may have contributed to scene type effects as explored first, but this seems to be less important than the orienting effects of the danger judgment task and the inherent complexities of comprehending spatial expressions, which are considered afterward.

One possible scene type effect could have been expected from the danger variation but was not confirmed in the results. Danger for some predator scenes differed within a sentence pair. If the danger judgment given in the study phase was remembered, it may have been used in the recognition test as episodic content. However, there is no evidence in the results supporting this consideration. Error rates for different-models sentences were similar for predator and peaceful scenes in all but one of the experiments employing the danger judgment task. Furthermore, only 10% of all recognition items were from predator-scene different-models sentence pairs with a change in precategorized danger.

The reoccurring pattern of unexpected scene type effects consisted of higher discriminability scores for predator scenes, especially same-model predator scenes, in experiments with the danger judgment task and extended encoding. This pattern suggests that the danger judgment task oriented participants differently toward predator and peaceful scenes. For peaceful scenes, it directed participants to evaluate spatial features, whereas predator scenes were more salient in the danger judgment task. For predator scenes, the evaluation of danger presumably led to more variable and richer encoding based on background knowledge.

The evaluation of predator-scene sentences may have involved the attribution of intentions to predator animals on the basis of the description of their activities. For example, the lion described as trotting toward the zebras next to the shrub may have been more strongly perceived as hostile than the lion described as trotting toward the shrub next to which the zebras graze. As this same-model sentence pair shows, these attributions could have differed even within same-model pairs. But this explanation can only be part of the story. A closer look at the results for single sentence groups revealed that the sentence groups responsible for scene type effects varied between experiments and that among the sentence groups responsible, there are those that do not describe intentional activities but instead describe static postures of predator animals.
Scene type-related variations seem more important in the interpretation of spatial expressions. For example, the orienting axis of reference objects may have induced variation mediated by scene type. The pheasant–puddle sentence group is a sentence group that consistently contributed to better recognition scores for predator-scene than it did for peaceful-scene new same-model sentences: A pheasant drinks in a puddle and a wolf/a lizard comes up to it (the puddle)/ it (the pheasant). In the predator-scene same-model pair, a wolf approaches the pheasant or the puddle. Although the pheasant and the puddle are in the same place, the two movement descriptions of this same-model pair may be interpreted differently because the puddle has no orienting axis or front side, but the pheasant has. Therefore, coming up to the pheasant will be interpreted as approaching the pheasant from its front. In contrast, coming up to the puddle could also be interpreted as approaching from behind the pheasant. This difference within the same-model pair is more important in the predator scene and may have been noticed with higher probability there. Similar scene type-mediated effects seem to have occurred involving the size of reference objects. Thus, the use of conceptual knowledge in interpreting spatial expressions (Morrow & Clark, 1989) presumably increased discriminability scores for same-model predator scenes.

The kind of conceptual knowledge that is used and its relative importance depends on the context and the comprehenders’ goals. This may explain why, in Experiment 1, without the danger judgment, some peaceful-scene same-model sentences were easier to discriminate. For example, the hen–bowl sentence group is among those that had lower error rates in the peaceful-scene same-model variant than they did in the predator-scene same-model variant in Experiment 1 but not in the experiments with the danger judgment. The hen–bowl sentence group reads: A hen pecks in a bowl and next to it (the bowl)/ it (the hen) crouches a marten/ a chick (a marten is a type of a weasel). The possible interpretations of the expressions next to the hen and next to the bowl overlap if the hen is located in the bowl; otherwise, they would not form a same-model pair. But next to the hen includes the space next to the hen in the bowl, whereas next to the bowl excludes this space. For the danger judgment task, presumably, different interpretations were explored to ensure that one that would constitute danger would not be missed. However, in Experiment 1, participants may have been satisfied with a single interpretation. This single interpretation may have differed between same-model pair sentences in the peaceful variant if the chicken crouching next to the hen was located in the bowl, rather than definitely next to the bowl, as it was in the other sentence. Thus, the peaceful-scene same-model pair sentences could have received discriminable interpretations. In contrast, the marten in the predator variant presumably had not been located in the bowl next to the hen. Please note that this explanation assumes that spatial representations sometimes were constructed for these peaceful-scene sentences in Experiment 1. It seems unlikely that no spatial situation models were constructed at all for peaceful-scene sentences in Experiment 1. But it was shown that the probability for construction was reduced, because the precondition of causal relevance was not fulfilled.

The two examples involving the orienting axis of reference objects and varying interpretations of next to illustrated possibilities for same-model pairs actually functioning as different-models pairs. This is a problem of the present paradigm that exploits ambiguities and overlaps in the interpretation of spatial expressions. Nonetheless, the increased variance did not prevent the model type effect, which was stable and confirmed the predicted importance of causal relevance. In future studies, even more effort should be invested in choosing sentence groups whose same-model pairs remain same-model pairs under a wide range of interpretations.

Conclusions

Despite complexities in the comprehension of spatial expressions and a variety of phenomena influencing recognition performance, the reported experiments, as a whole, confirmed the importance of causal relevance as a precondition for the spontaneous construction of spatial situation models. Spontaneous construction of spatial situation models and, presumably, other spontaneous elaborative activity during text comprehension is more probable with texts that get “minimalists” to work (Foerstch & Gernsbacher, 1994, p. 271). In studies of naturalistic text comprehension, experimental materials must not be too dry, schematic, and boring, especially if spontaneous activity is being studied. Grice’s (1975) maxim of relevance is not suspended for spatial descriptions given in experiments. Therefore, the last things you need in matters of this kind are sleeping turtles.

References

Dutke, S. (1996). Generic and generative knowledge: Memory schemata in
the construction of mental models. In W. Battmann & S. Dutke (Eds.), Processes of the molar regulation of behavior (pp. 35–54). Scottsdale, AZ:当代心理学家.


In G. Rickheit & C. Habel (Eds.), Mental models in discourse processing and reasoning, Advances in Psychology 128 (pp. 93–112). Amsterdam: Elsevier.


(Appendix follows)
German Text Used in the Experiments [With English Translation]

Experimental sentence groups

   [Two zebras graze next to a shrub and a lion/an antelope trots towards it/them.]
   [Two zebras move away from a shrub and a lion/an antelope trots towards it/them.]
   [A pheasant drinks in a puddle and a wolf/a lizard comes up to it (the puddle)/it (the pheasant).]
   [A pheasant runs away from a puddle and a wolf/a lizard comes up to it (the puddle)/it (the pheasant).]
5. Ein Huhn pickt in einer Schüssel und bei ihr/ihm duckt sich ein Marder/ein Küken.
   [A hen pecks in a bowl and next to it (the bowl)/it (the hen) crouches a marten/a chick.]
   [A cat/a pigeon sits on a cage and above it (the cage)/it (the cat/pigeon) flutters a budgerigar.]
   [A seal comes up from an ice hole and above it (the ice hole)/it (the seal) stands a polar bear/a seagull.]
   [Two rattlesnakes/earthworms lie in a crevice and above it/them sniff around a rabbit.]
   [A squirrel climbs up a tree and the dogs/the geese surround it (the tree)/it (the squirrel).]
   [An owl/a blackbird lands on a peg and mice sit around it (the peg)/it (the owl/blackbird).]
    [Two hares look out of a burrow and in front of it/them stands a fox/a roe.]
    [A sheep steps out of a cave and in front of it (the cave)/it (the sheep) stands a bear/a goat.]

Filler sentence pairs

F1. Ein Luchs steht hinter einem Pferd, als es ausschlägt.
   [A lynx is standing behind a horse as it kicks out.]
F2. Ein Elefant läuft auf einen Leoparden zu.
   [An elephant trots towards a leopard.]
New Editor Appointed for Journal of Occupational Health Psychology

The American Psychological Association announces the appointment of Lois E. Tetrick, PhD, as editor of Journal of Occupational Health Psychology for a 5-year term (2006–2010).

As of January 1, 2005, manuscripts should be submitted electronically via the journal’s Manuscript Submission Portal (www.apa.org/journals/ocp.html). Authors who are unable to do so should correspond with the editor’s office about alternatives:

Lois E. Tetrick, PhD
Incoming Editor, JOHP
George Mason University
Department of Psychology, MSN, 3F5
4400 University Drive, Fairfax, VA 22030

Manuscript submission patterns make the precise date of completion of the 2005 volume uncertain. The current editor, Julian Barling, PhD, will receive and consider manuscripts through December 31, 2004. Should the 2005 volume be completed before that date, manuscripts will be redirected to the new editor for consideration in the 2006 volume.

Appendix (continued)

Filler sentence pairs (continued)

F3. Eine Kuh sieht ein Wiesel vorbeilaufen.
   [A cow watches a weasel pass by.]
   Eine Kuh sieht, wie ein Wiesel wegläuft.
   [A cow watches a weasel running away.]
F4. Eine Hyäne macht einen Bogen um einen Büffel.
   [An hyena by-passes a buffalo.]
   Eine Hyäne hält Abstand zu einem Büffel.
   [An hyena keeps its distance from a buffalo.]
F5. Ein Esel tritt auf einen Erdhügel, in dem ein Maulwurf gräbt.
   [A donkey steps on a mound, in which a mole digs.]
   Ein Esel läuft über einen Erdhügel, in dem ein Maulwurf gräbt.
   [A donkey trots over a mound, in which a mole digs.]
   [A beaver fells a tree, under which hedgehogs eat.]
   Ein Biber fällt einen Baum, unter dem Igel Nahrung suchen.
   [A beaver fells a tree, under which hedgehogs search for food.]

Note. F = filler.