Memory & Cognition
2006, 34 (4), 817-825

Memory for the process of constructing an integrated mental model

STEPHEN J. PAYNE
Cardiff University, Cardiff, Wales

and

THOM BAGULEY
Nottingham Trent University, Nottingham, England

We report two experiments in which people read descriptions of integrated spatial configurations, together with comparable descriptions that did not describe integrated spatial configurations. The integrated spatial descriptions, but not the comparable descriptions, thus supported the construction of a coherent mental model. In Experiment 1, each sentence of the comparable descriptions described the spatial relation between two objects that were not mentioned elsewhere in the description. In Experiment 2, the comparable descriptions were nonspatial, having been constructed by replacing the spatial relations with nonspatial relations. In both experiments, participants were given a surprise recognition test in which they had to identify each of the original descriptions—of both integrated spatial configurations and nonspatial configurations—among a set of distractors. When the sentences in the original description were reordered (and participants were instructed to ignore sentence order), recognition memory was reliably depressed, but only for the integrated spatial descriptions. Reordering descriptions does not change their propositional content, nor does it change the described situation; however, it does change the process of constructing a mental model of that situation. These findings thus suggest that memory for the descriptions retains a trace of the process of constructing an integrated mental model and that reordering the sentences disrupts this memory because the reordering reduces the similarity of the processing of the descriptions at recognition.

In previous work on the comprehension and memory of simple spatial and temporal descriptions (Baguley & Payne, 1999, 2000; Payne, 1993), we have defended the suggestion that people comprehend these descriptions by constructing mental models of the premises and proposed that their memory of the descriptions contains a trace of the processing steps of model construction—an episodic construction trace.

This suggestion instantiates a processing view of human memory in a specific cognitive domain—namely, understanding and remembering text. The traditional view of memory for comprehended text distinguishes memory for the text itself (including surface form and propositional content) from memory for the situation described by the text (e.g., Bower & Morrow, 1990; Gernsbacher, 1990; van Dijk and Kintsch, 1983). Our suggestion goes beyond this in proposing that the episodic record of the process of constructing a situation model is also remembered.

Our claim also relates to the important transfer-appropriate processing tradition (e.g., Morris, Bransford, & Franks, 1977). The essential tenet of this tradition is that the processes that are performed during encoding influence memory in more particular ways than a correlation between "depth" and robustness of the memory (see Craik & Lockhart, 1972). The essential tenet, however, can be interpreted in relatively weaker or stronger versions of the role of mental processes in memory. A weak version would contend that all memory traces have multiple features, and that the processes at encoding merely emphasize some of these features at the expense of others. For example, judging the semantics of a word will emphasize the semantic features in its trace, whereas judging it for rhyme will instead emphasize the trace's phonetic features. At the other extreme, a strong version would contend that the memory trace is simply the process of encoding (e.g., Crowder, 1993; Kolers, 1973; Lansdale, 2005). Under this hypothesis, recognition would result from the awareness that a mental process or parts of that process were being repeated, and thus recognition would not require any explicit comparison. A fundamental distinction between weaker and stronger versions of transfer-appropriate processing is made between theories that propose that merely the end product of processing is retained and theories that propose that intermediate processing is also retained. In between the two extremes are myriad alternatives, which

---

This research was partially funded by the Economic and Social Sciences Research Council, U.K., under Grant R000235641. Correspondence should be addressed to S. Payne, School of Informatics, University of Manchester, P.O. Box 88, Sackville Street, Manchester M60 1QD, England (e-mail: stephen.payne@manchester.ac.uk).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

make differing assumptions about the nature of cognitive architecture in which processing arises.

Our position, which we will support with empirical evidence, is relatively strong. We claim that it is not merely the end product of comprehension processes that is remembered, but also the vital constructive steps that constitute that process. A nice distinction can be made here between the claim that these vital constructive steps are encoded continuously and the claim that discrete intermediate steps are recorded. We are agnostic on this point (and indeed we are unsure that it is possible to distinguish empirically such a continuous view of memory from one in which intermediate results or goals of processing are encoded). Under both the discrete and continuous views, the more well specified the encoding processes, the more specific the empirical consequences that follow. We adopt Johnson-Laird's (1983) proposal for how mental models of simple descriptions might be constructed, and we predict that the form of each step in this process will affect memory. We describe our proposal in a simple propositional notation (a list of steps), but it is the content of memory that we are predicting, not its form.

For example, consider a person who has read and comprehended this simple spatial description:

> The coke is below the scotch.
>
> The coke is above the vodka.
>
> The coke is to the left of the wine.
>
> The beer is above the wine.

Following Johnson-Laird and colleagues (e.g., Byrne & Johnson-Laird, 1989; Johnson-Laird, 1983; Mani & Johnson-Laird, 1982), we suppose that the reader constructs an integrated mental model of the layout of the drinks. We assume that readers typically construct this model incrementally, adding new objects as they are encountered to ones already in the model. These processing steps can be described as follows, using the list notation described by Payne (1993):

> [start [coke scotch below]]
>
> [vodka coke below]
>
> [wine coke right]
>
> [beer wine above]

In this notation, each line denotes an operation in which the first object is newly added to the model in the specified location relative to the second object (which is already in the model). The beginning of model construction needs to be distinguished using the argument "start." Note that this putative construction trace is neutral concerning the order of mention within the individual sentences. For example, the trace would be just the same if the second sentence of the description had been

> The vodka is below the coke.

However, the trace is critically dependent on the order of the sentences within the description, because this determines the order in which objects are added to the model.

In the original formulation of the episodic construction trace, it was suggested that this trace may be the only representation of a description that remains in long-term memory (Payne, 1993). However, our more recent experiments have convinced us that it coexists with propositional representations and with a partially-remembered mental model (Baguley & Payne, 2000). Because the suggestion of remembered propositions and mental models is far from controversial (e.g., Fletcher, 1994; van Dijk & Kintsch, 1983), it is the idea that a construction trace might be stored and might influence subsequent performance that demands the most defense and deserves the most empirical exploration.

In published studies, we have provided several different strands of evidence for the episodic construction trace. Using a paradigm introduced by Mani and Johnson-Laird (1982) in which participants must distinguish an original description from various foils, including one that describes the same configuration of objects, we have shown that the similarity, in terms of the episodic construction trace, between this "inferable" foil and the original affects recognition performance (Baguley & Payne, 2000; Payne, 1993).

Converging evidence comes from a single-sentence recognition task. Baguley and Payne (1999) showed that people, when trying to recognize single sentences from descriptions like the one above, are significantly more likely to judge a sentence that is consistent with the construction trace as having already been read, whether or not the form of this sentence is indeed identical to the sentence they had read. For example, in the description above, participants are more likely to accept as "old" the sentence *The vodka is below the coke* than the sentence that they actually did read (*The coke is above the vodka*).

Finally, as suggested above, we have reported that recognition is depressed when the sentences of the original descriptions are reordered at test, even though participants are explicitly instructed to ignore sentence order in making recognition judgments (Baguley, 1994; Baguley & Payne, 2000; Payne, 1993). This is perhaps the most intuitively compelling of all the predictions that we have tested. Reordering the sentences of a description at test does not change the sentences themselves; therefore the effect cannot be explained in terms of memory for the sentences or their propositional form. Nor does reordering alter the situation presented by the description, so the effect also cannot be explained in terms of memory for the structure of a mental model. Instead, we argue that reordering influences recognition scores by changing the order in which a mental model is constructed, which in turn influences the episodic record of the operations used to build up the mental model.

Is any other aspect of the description altered by reordering? Payne (1993) noted that "some syntactic or phonological encoding" may be sensitive to reordering. Such a proposal would be consistent with the view that in-context recognition tests are more sensitive to memory for surface form than are sentence-based tests (see, e.g., Fletcher, 1992). However, demonstrating memory for surface form is notoriously difficult (Fletcher, 1994), whereas the reordering effect in our experiments is relatively robust and easy to replicate.

Nevertheless, it would be more compelling to demonstrate that the reordering effect is conditional on a reduction in trace overlap between the original description and its reordered counterpart in the recognition test. This requires finding a case where the episodic construction trace account predicts that reordering sentences would *not* reduce trace overlap. To develop such a case, let us reconsider why reordering does reduce trace overlap for the descriptions we have used. Reordering alters the episodic construction trace because it changes the order in which objects are entered into a mental model. The trace records which objects are new and which are already in the model as each sentence of the description is processed. Thus the reordering effect is conditional on participants' reading sentences that refer to objects or events mentioned in previous sentences. If we present participants with sentences that do not refer to objects or events mentioned in earlier sentences, the episodic construction trace hypothesis predicts no effect of reordering on recognition scores.

For example, consider again the determinate spatial description above. If we substitute a new drink for each occurrence of an old drink, we can create the following description:

The coke is below the scotch.
The *cider* is above the vodka.
The *lemonade* is to the left of the wine.
The beer is above the *brandy*.

The resulting description does not describe a single situation but rather four unconnected states of affairs (or four separate models). We term this type of description a *multiple-model* description, as distinguished from a *one-model* description in which the situations described are connected. Reordering one-model descriptions lowers the trace overlap between the reordered and the original descriptions, as demonstrated in this paper and by Payne (1993). Let us assume that when participants read a multiple-model description, they build, in sequence, four distinct mental models. In this case the episodic construction trace for the multiple-model description would be

[start [coke scotch below]]
[start [cider vodka above]]
[start [lemonade wine left]]
[start [beer brandy above]]

Reordering the sentences would affect the order of sublists within this trace (which is assumed to be irrelevant in recognition judgments) but would have no effect on the sublists themselves. Therefore, the episodic construction trace hypothesis predicts that reordering a multiple-model description should have no effect on recognition memory. It could be objected that participants may not bother to build up mental models for each individual sentence. However, this does not alter our prediction, since in this case no episodic construction trace would be encoded. If no episodic construction trace is present, we also predict that reordering would have no effect on recognition memory, which relies on propositions or retention of surface form. (Putting this argument to one side for the sake of simplicity, we will refer to such descriptions as *multiple-model descriptions*.)

The implications of any empirical support for the episodic construction trace hypothesis are twofold. First, such support would furnish converging evidence for mental models. This evidence relies on assumptions about how the models are constructed and their influence on memory-for-processing, rather than on models' static properties (such as their complexity or their analog form), or the way the models are consulted during inference. An existing literature relates the key predictions of mental model theory to recognition memory (e.g., Glenberg, Meyer, & Lindem, 1987; Radvansky & Zacks, 1991; Zwaan, 1996), but the processing angle of our hypothesis makes a novel contribution.

Second, as discussed above, empirical support for the episodic construction trace hypothesis is evidence for a theory of memory in which processes at encoding are strongly considered to be a principal determinant of what is remembered. This conclusion falls short of one's taking a stance with regard to how trace fits into cognitive architecture; for example, one might imagine implementing the key properties of the episodic construction trace in an instance architecture (Hintzman, 1986; Lansdale, 2005) or in a more traditional architecture such as ACT-R (Miles, Payne, & Baguley, 1998). The issue will be returned to below.

EXPERIMENT 1

In this experiment, we investigated the effect of reordering one-model and multiple-model descriptions on recognition memory. This is a strong test of the episodic construction trace hypothesis. If the reordering effect is a consequence of the removal of context or order information (e.g., in some representation of surface form), then recognition memory for both one-model and multiple-model descriptions should be impaired by reordering. If the reordering effect is due to disruption of the episodic construction trace, as we propose, then recognition memory should be impaired only for the one-model descriptions.

Method

Participants. Sixty people took part in the experiment. Participants were recruited from Cardiff University and were paid £4.00 or given course credit on completion of the experiment.

Materials. Descriptions for the learning phase (Phase 1) were based on materials used by Mani and Johnson-Laird (1982), Payne (1993), and Baguley and Payne (2000). Four determinate one-model descriptions were used (excluding the two practice descriptions); each description contained four sentences connecting five objects. Each sentence consisted of two concrete nouns linked by a spatial relation (*to the left of*, *to the right of*, *in front of*, or *behind*). Referential continuity was maintained in the descriptions by introducing one new object in each sentence after the first.

A corresponding set of multiple-model descriptions was also created. This set was identical to the one-model descriptions, except that each recurrence of an old object (an object mentioned in a previous sentence) was replaced with a new object from the same category (see the introduction to this experiment for an example of a one-model description and the multiple-model description generated from it, and the Appendix for a further example).

Each description was headed by a category to which all of the objects belonged. The categories were chosen to minimize confusion among descriptions. Categories were randomly allocated to either one-model or multiple-model descriptions, so that each participant received a random combination of category type and descriptions. The eight categories used were animals, birds, clothing, drinks, fruit, musical instruments, gemstones, and vegetables, and each category contained eight objects. These objects were randomly assigned to roles in the description (for the one-model description, five of the eight objects were chosen randomly from the list each time materials were generated for a new participant).

For each description presented in the learning phase, four related descriptions were presented in the recognition test. The recipe for constructing the four related descriptions was identical for the one-model and the multiple-model descriptions. An inferable description was created by leaving two sentences unchanged and inverting the remaining two sentences (e.g., *The vest is below the kilt* would become *The kilt is above the vest*). A foil based on the original description was constructed by swapping the roles of two objects in the original description, and a foil based on the inferable description was constructed by swapping the roles of the same two objects in the inferable description. The selection of these two objects was constrained to leave exactly two sentences of each foil unchanged from the original description. Thus the two foils described the same situation as each other, but, like the inferable description, they each shared two sentences with the original description.

Half of the one-model and half of the multiple-model descriptions, chosen at random, had their original descriptions reordered at test for each participant. Reordered versions of the one-model descriptions were constructed to have a trace overlap of 1 (i.e., only one sublist was common to the construction traces for the original and reordered descriptions). The multiple-model descriptions were reordered in exactly the same way. However, as noted in the introduction, reordering a multiple-model description has no effect on trace overlap. Trace overlap for the remaining recognition test items from the one-model descriptions was 2 for the inferable descriptions and 1 for each foil. Trace overlap for the remaining recognition test items from the multiple-model descriptions was based on the number of shared sentences (in this case, 2 each for the inferable and for both foils).

Procedure. The experiment was divided into two phases: a learning phase and a surprise recognition test. Before starting the main experiment, the participants were given practice trials with a one-model and a multiple-model description. During the learning phase, the participants read each description and were then shown a pair of objects from one of the sentences in that description. They had to decide whether the objects were displayed in positions consistent with those in the description they had just read, and indicate their decision by selecting "true" or "false" on the computer screen. The entire set of descriptions was read four times, with a relation verification task after each presentation of a description.

Because the hypothesis was concerned with verbatim rather than gist memory, the Phase 2 recognition test was a one-of-four forced-choice recognition, rather than the ranking test used in earlier work (Baguley & Payne, 2000; Mani & Johnson-Laird, 1982; Payne, 1993). The participants were presented with the four alternative descriptions on the computer screen and were simply asked to select the description that they thought was most similar to the original description they had read. They were told that each trial included the original description, but that the sentences in this description might be in a different order. They were instructed to ignore the order of the sentences within a description when making a response.

Results

Learning. Performance on the relation verification task was slightly better for the one-model (87%) than for the multiple-model (84%) condition, though this difference was only marginally significant [$t(59) = 1.94$, $SE = 1.50$, $p = .057$]. It may be that the multiple-model condition was slightly harder because the description contained relationships among eight objects rather than five. Three participants performed at chance (exactly 50%) on the relation verification test. Because these participants showed no evidence of comprehension in Phase 1, their data were excluded from subsequent analyses (although including these data would not alter the pattern of significance obtained).

Recognition. Recognition was scored if participants chose the original description, in its original or reordered form. The mean and standard deviation recognition scores for the original and reordered one-model and multiple-model descriptions are shown in Table 1.

ANOVA was carried out on the recognition scores with description type (one-model vs. multiple-model) and reordering (original vs. reordered) as factors. There was no significant effect of description type ($F < 1$). A significant main effect of reordering [$F(1,56) = 7.01$, $MS_e = 1,141$, $p < .05$] and a significant interaction between reordering and description type [$F(1,56) = 4.63$, $MS_e = 856$, $p < .05$] were obtained. Simple main effects were computed to explore the interaction further. As predicted by the episodic construction trace hypothesis, reordering significantly impaired recognition of the one-model descriptions [$F(1,56) = 9.678$, $MS_e = 1,199$, $p < .05$], but had no significant effect on the multiple-model descriptions ($F < 1$).

An analysis was also performed on the type of errors in the recognition task. For both the one-model and the multiple-model descriptions, the inferable item was chosen slightly more often than the chance rate of 1 in 3 ($M = 44\%$ for the one-model descriptions and $M = 46\%$ for the multiple-model descriptions). This difference was not statistically significant ($t < 1$). Aggregating over description type, the inferable was chosen 47% of the time [$t(54) = 3.17$, $SE = 4.43$, $p < .05$]. This preference for the inferable description suggests that participants retained information about the spatial relationships in both one- and multiple-model conditions. (The proportion of the remaining errors

Table 1
Recognition Scores in Experiment 1, Means and Standard Deviations

Type	Description	*M*	*SD*
One model	Original	66	34.0
	Reordered	46	37.8
Multiple model	Ordered	61	33.8
	Reordered	58	30.8

was similar between foil type and description type and not analyzed further.)

Discussion

Reordering a description at test significantly impaired memory for one-model descriptions but not for multiple-model descriptions. This outcome suggests that reordering a description does not influence memory by disrupting a surface form representation of the order of the sentences. Rather, reordering disrupts recognition memory only when it also reduces the episodic construction trace overlap between a description at learning and at test. This is consistent with the independent effects of reordering and inferable trace overlap observed by Baguley and Payne (2000, Experiment 3) and with the influence of inferable trace overlap on recognition memory reported by Payne (1993).

However, a new explanation of the reordering effect was suggested by this experiment. Perhaps the effect relies simply on spreading activation, during recognition, between propositions that are linked in a network because of shared arguments. According to this idea, study materials are represented in memory as a network of associated propositions. As one of these propositions is read during recognition, its memory trace becomes activated, and activation spreads to adjacent propositions in the network. If the next proposition read is one of these adjacent propositions (as it would be in original but not reordered descriptions), recognition is boosted. This effect might be removed for the multiple-model descriptions in which the propositions share no arguments. One might expect that such propositions would be more weakly associated in memory and would therefore serve less well as retrieval cues for one another.

We believe that this theory might be attacked, particularly because it relies on associations between adjacent propositions being very much stronger than other associations within the network. In fact, there is evidence that the associative effects of propositional adjacency (adjacency of occurrence in the text) can be overwhelmed by what we might term "narrative adjacency," whether spatial (one object being spatially close to another; Glenberg, Meyer, & Lindem, 1987) or temporal (one event being only a moment rather than an hour after another; Zwaan, 1996). Nevertheless, the supposition of a network representation in memory fits well with contemporary theories (e.g., ACT-R; Anderson, 1993) and cannot be ruled out on the basis of published data or Experiment 1.

EXPERIMENT 2

In this experiment, we replaced the multiple-model condition of Experiment 1 with another kind of description, which, like the single-model spatial descriptions, allowed arguments to be shared between adjacent propositions but, unlike the spatial descriptions, did not allow a single mental model to be constructed. We accomplished this by replacing spatial relations with nonspatial relations so that the two dimensions of variation could not readily be integrated.

During a pilot within-subjects version of this experiment, some participants appeared to spatialize nonspatial relations such as "darker than" and "lighter than." Therefore, we chose dimensions (wetter/dryer and harder/softer) that we felt would not be spontaneously spatialized. Although it is true that any relations can in principle be spatialized, to integrate independent relations into a spatialized model would require a participant to adopt consistent yet arbitrary assumptions about the representation of each relation and the metarelation between the relations (e.g., their orthogonality). To be effective in the experimental task, such assumptions would have to be made quickly and consistently by each participant and be sufficiently easy to enact that they would not draw processing resources from the relatively effortful process of reading and understanding a spatial description. Furthermore, there is evidence that rather minor manipulations of materials can inhibit participants' ability or willingness to form integrated models. For example, disrupting the referential continuity of the text propositions has such an effect (Ehrlich & Johnson-Laird, 1982), as do similar but semantically distinct set inclusion premises which tend not to be integrated (Favrel & Barrouillet, 2000). To further discourage spatialization of nonspatial relations, we moved to a between-subjects design in which participants were shown either spatial or nonspatial descriptions exclusively, which minimized the possibility that participants would exploit the similarity between spatial and nonspatial forms in order to integrate nonspatial descriptions by analogy.

According to a propositional network account of reordering effects, recognition of such nonspatial descriptions would be disrupted by reordering to an extent equivalent to that for the spatial descriptions, whereas our episodic construction trace account predicts that reordering would have a smaller effect on recognition of these descriptions.

Method

Participants. Fifty people took part in the experiment. Participants were recruited from Cardiff University or Loughborough University and were paid £4.00 or given course credit on completion of the experiment.

Materials. A set of eight one-model spatial descriptions was constructed on the basis of the recipes for determinate descriptions in Experiment 1. A corresponding set of eight nonspatial, multiple-model, mixed-relation descriptions was also created. This set was identical to the one-model descriptions, except that spatial relation terms were systematically replaced by two nonspatial relation terms (harder/softer and wetter/dryer). For example, *to the left of* was replaced by *harder than*, and *to the right of* by *softer than*. See the Appendix for sample materials.

The recipe for constructing the foil items in the recognition test was identical for spatial and nonspatial descriptions and was amended slightly from Experiment 1 to reduce the sentence and trace overlap of some items with the original descriptions. Inferable descriptions were created by leaving one sentence unchanged and inverting the remaining three sentences. The foil based on the original description was constructed by swapping the roles of two objects in the original description. The foil based on the inferable description was constructed by swapping the roles of the same two objects in the inferable description. The selection of these two objects was constrained to minimize sentence and trace overlap with the original description. Trace and sentence overlap was held at one for the foil based on the inferable description and at two for the foil based on the original description. As in Experiment 1, half of the descriptions of each type, chosen at random, had their original descriptions reordered at test for each participant. Reordered versions of the one-model descriptions were constructed to have a trace overlap of one. The nonspatial reordered descriptions (which were reordered in the same way as the spatial descriptions) had a trace overlap of four.

Procedure. The procedure was identical to that in Experiment 1, except in three respects. First, participants were randomly assigned to either a spatial condition or a nonspatial condition between subjects. To compensate for the reduced statistical power of the between-subjects manipulation, the number of descriptions of each type was increased to eight (therefore reducing measurement error).

Second, a sentence verification task was used in place of the relation verification task to accommodate the materials for the nonspatial condition. After reading each description, participants were shown a sentence based on one of the four sentences they had read. Half of the trials presented an inverted but consistent sentence; for example, if the participant had read "The apple is to the left of the banana," he/she might be shown "The banana is to the right of the apple," in which case the correct response was "true." The other half of the trials reversed the object order but retained the relation; for example, "The banana is to the left of the apple" and the correct response was "false."

Third, the number of Phase 1 trials was reduced for the nonspatial descriptions. Two Phase 1 trials were used in the nonspatial condition and four in the spatial condition. A pilot study for this experiment used the same number of Phase 1 trials for spatial and nonspatial descriptions. Although the pattern of significance was the same as in the data reported here, there was a suggestion of a ceiling effect for the nonspatial materials in Phase 2, suggesting that the task was slightly too easy. As such a ceiling effect could mask a reordering effect, the study was redesigned with fewer Phase 1 trials to improve the opportunity to detect any reordering effect.

Results

Learning. Performance on the sentence verification task was slightly better for the spatial (80%) than for the nonspatial (74%) condition, though this difference was not significant [$t(48) = 1.63$, $SE = 3.97$, $p = .11$]. Four participants were performing at or slightly below chance (50%) on the sentence verification test. As these participants showed no evidence of comprehension in Phase 1, their data were excluded from subsequent analyses (including these data would not alter the pattern of significance obtained).

Recognition. Recognition was scored if participants chose the original description, in its original or reordered form. The mean and standard deviation recognition scores for the original and reordered spatial and nonspatial descriptions are shown in Table 2.

ANOVA was carried out on the recognition scores with description type (spatial vs. nonspatial) and reordering (original vs. reordered) as factors. There were significant main effects of both description type [$F(1,44) = 5.09$, $MS_e = 707$, $p < .05$], and order [$F(1,44) = 12.40$, $MS_e = 461$, $p < .05$]. A significant interaction between reordering and description type was obtained [$F(1,44) = 5.32$, $MS_e = 461$, $p < .05$]. Simple main effects were computed to explore the interaction further. As predicted by the episodic construction trace hypothesis, reordering significantly impaired recognition of the spatial descriptions [$F(1,22) = 12.83$, $MS_e = 610$, $p < .05$], but had no significant effect on the nonspatial descriptions [$F(1,22) = 1.09$, $MS_e = 311$, $p = .31$].

Table 2
Recognition Scores in Experiment 2, Means and Standard Deviations

Type	Description	*M*	*SD*
Spatial	Original	70	26.1
	Reordered	43	25.3
Nonspatial	Original	72	21.7
	Reordered	66	23.4

As in Experiment 1, an analysis of errors for the two description types was also performed. The pattern of errors varied considerably between conditions and is summarized in Table 3.

A higher proportion of inferable descriptions were chosen by participants in the spatial condition than in the nonspatial condition [$t(41) = 2.03$, $SE = 9.79$, $p < .05$]. Conversely, a higher proportion of foils based on the original description were selected by participants in the nonspatial condition than in the spatial condition [$t(41) = 2.69$, $SE = 8.56$, $p < .05$]. In both cases, the foil based on the inferable description was the least common error response; this is not surprising, since in addition to sharing only one sentence with the original description, it also describes objects with a different spatial configuration than that of the original.

This pattern of errors is consistent with the view that participants in the spatial conditions constructed and remembered a spatial mental model. Participants in the nonspatial conditions, as predicted, seemed to find it hard to form an integrated representation of the sentences (and therefore produced fewer errors that preserved the relationships among the objects in the description). One interpretation of their pattern of errors is that on trials where they could not identify the original description, they were able to avoid the foil based on the inferable description and to guess at the remaining items. This would explain the very similar proportion of errors where the inferable description and the foil based on the original description were selected.

Discussion

As in Experiment 1, reordering a description between study and test impaired recognition memory of spatial descriptions. No such reordering effect was found for nonspatial descriptions that were identical in propositional form. In this experiment, unlike in Experiment 1, the argument overlap between adjacent propositions was identical in the two conditions and therefore could not explain the difference in reordering effects.

The absence of a reordering effect for nonspatial descriptions supports our hypothesis that such effects are caused by a disruption to the process of constructing an integrated mental model of a description. When participants do not construct such a model, the effect of reordering is substantially reduced. The pattern of recognition errors further supports our a priori assumption that people do form integrated

Table 3
Pattern of Errors for Spatial and Nonspatial Descriptions in Experiment 2

	Error Type					
	Inferable		Foil O		Foil I	
Description Type	*M*	*SD*	*M*	*SD*	*M*	*SD*
Spatial	68	35.8	22	25.9	11	25.6
Nonspatial	48	31.6	45	34.2	8	15.0

Note—*Foil O* and *Foil I* refer to the foil descriptions based on the original description and on the inferable description, respectively.

models of the spatial descriptions but cannot or do not construct such models for the nonspatial versions.

GENERAL DISCUSSION

In two experiments, we replicated a large disruptive effect on recognition of spatial descriptions reordered between study and test. Reordered spatial descriptions contain the same propositions and describe the same situation as do their original versions; hence this reordering effect cannot readily be explained by memory for text or memory for mental models. In earlier work (Baguley & Payne, 2000; Payne, 1993), we have proposed that the reordering effect is caused by a disruption to the process of constructing a mental model from the description.

In the first experiment, we pitted this explanation against a nonspecific alternative that proposed that memory for the descriptions depended on preserved adjacency of sentences. We showed that the reordering effect is significantly reduced when sentences in a description cannot be integrated into a single mental model because they do not have arguments in common. This suggests that the reordering effect depends on the process of constructing a single mental model, in keeping with our account.

In the second experiment, we pitted our explanation against an alternative account that proposed the spreading of memory trace activation to adjacent propositions in a network. According to this alternative account, the reduction of the reordering effect in Experiment 1 may have been due to propositions within a description being less strongly connected when they did not share arguments. In Experiment 2, the propositional structure of spatial and nonspatial descriptions was identical in this sense, yet the reordering effect was substantially reduced in the nonspatial case.

Taken together, these experiments offer compelling support for our account of the reordering effect on recognition memory for spatial descriptions. The results of our experiments add to the body of evidence that people who read spatial descriptions construct mental models of the situations described and that their memory incorporates a trace of the construction process itself.

Finally, we would argue that the rather well-developed notion of processing that is inherent in the mental models account of comprehension of the kind of simple descriptions that we have studied has allowed us to offer support for the general idea that memory preserves cognitive processes as well as cognitive objects. This is consistent with some important conjectures about the nature of human memory (see, e.g., Kolers, 1973; Lansdale, 2005).

The Kolers (1973) conjecture, as we interpret it, is that memory for an object or an event is more accurately understood as the processes that encode and reflect on the external stimulus, rather than some record or engram of the stimulus itself. Under this proposal, which to our knowledge has never been implemented computationally, recognition, for example, would be a judgment that a mental process or part of a mental process is being repeated, rather than a comparison between a trace and a percept. The episodic construction trace hypothesis, and especially the evidence reported in this paper, are consistent with this radical position in that changing the comprehension process between study items and recognition items suppresses recognition judgments. However, the way in which we have expressed the episodic construction trace is somewhat less radical, and the data are fully consistent with just such a less radical alternative—namely, that intermediate steps during the comprehension process are stored as memory traces, as is the resulting mental model. Although less radical than Kolers's suggestion, this hypothesis is interesting not merely in the particular domain of text comprehension and mental model construction, but also for human memory more generally. It embraces what Lansdale (2005) calls the *acropetal memory*, in which representations emanating from all stages of the processing of an individual stimulus can be registered in memory.

Acropetal memory could be implemented by several cognitive architectures, and our proposals are intended to be neutral in this respect. For example, as noted by Radvansky (personal communication) and as earlier alluded to by Baguley (1994), one might consider the process of constructing a mental model from the descriptions used in our experiments as an incremental building, one sentence at a time, of bigger and bigger models. If every partial model is stored in memory as an instance, then relations mentioned earlier will occur in more remembered mental models, and reordered descriptions will disrupt any recognition processes that rely on this correlation. Whether such a sketch could be fully implemented is unclear (it may rely on recognition's somehow "knowing" that more models imply earlier processed relations), but the general scheme seems plausible. Our key argument, however, is that such a scheme would be an implementation of our theory, not a competitor. We have used a propositional trace notation to denote an abstract, functional claim about what is remembered, not a claim about implementation. Indeed, in earlier work we have shown that the episodic construction trace can be implemented in the ACT-R architecture, using memory for the goals that need to be stacked and executed to construct mental models based on spatial coordinates (Miles, Payne, & Baguley, 1998). This model showed that the construction of spatial mental models, the analogue form of such models, and memory for the construction process could all be represented within a relatively conventional cognitive architecture, in terms of goals, production rules, and a feature-based declarative memory.

Like the theory of mental models on which it is predicated (Johnson-Laird, 1983), the episodic construction trace hypothesis is about human cognition at the functional level. The experiments in this article support hypotheses in relation to both the understanding and the memory of simple spatial descriptions. When people read such descriptions, they construct analogue models of the described situations, and remember so doing. When their memory is tested, the process of constructing the model, as well as the constructed model, influence their memory performance.

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought*. Hove, U.K.: Erlbaum.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Baguley, T. S. (1994). *The representation of spatial mental models in long-term memory.* Unpublished PhD thesis, The Open University, Milton Keynes, U.K.

Baguley, T. S., & Payne, S. J. (1999). Recognition memory for sentences from spatial descriptions: A test of the episodic construction trace hypothesis. *Memory & Cognition*, **27**, 962-973.

Baguley, T. S., & Payne, S. J. (2000). Long-term memory for spatial and temporal mental models includes construction processes and model structure. *Quarterly Journal of Experimental Psychology*, **53A**, 479-512.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, **247**, 44-48.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory & Language*, **28**, 564-575.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, **11**, 671-684.

Crowder, R. G. (1993). Systems and principles in memory theory: Another critique of pure memory. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 139-161). Hove, U.K.: Erlbaum.

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning & Verbal Behavior*, **21**, 296-306.

Favrel, J., & Barrouillet, P. (2000). On the relation between representations constructed from text comprehension and transitive inference production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 187-203.

Fletcher, C. R. (1992). Assessing recognition memory for surface forms in discourse: A methodological note. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 199-203.

Fletcher, C. R. (1994). Levels of representation in memory for discourse. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589-607). San Diego: Academic Press.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language*, **26**, 69-83.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

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

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition*, **1**, 347-355.

Lansdale, M. W. (2005). When nothing is "off the record": Exploring the theoretical implications of the continuous recording of cognitive process in memory. *Memory*, **13**, 31-50.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, **10**, 181-187.

Miles, G. E., Payne, S. J., & Baguley, T. S. (1998). A production system model of memory for spatial descriptions. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of Cognitive Science Society* (pp. 699-705). Mahwah, NJ: Erlbaum.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 519-533.

Payne, S. J. (1993). Memory for mental models of spatial descriptions: An episodic-construction-trace hypothesis. *Memory & Cognition*, **21**, 591-603.

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 940-953.

Roediger, H. L., III, Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: The impetus from the levels-of-processing framework. *Memory*, **10**, 319-332.

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

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1196-1207.

APPENDIX

Sample spatial (one-model) description from Experiments 1 and 2:

The potato is to the left of the carrot.
The carrot is above the asparagus.
The lettuce is to the left of the asparagus.
The cabbage is below the asparagus.

Sample spatial (multiple-model) description from Experiment 1:

The violin is to the left of the guitar.
The saxophone is to the left of the trombone.
The harp is to the left of the accordion.
The flute is below the cello.

Sample nonspatial description from Experiment 2:

The kilt is softer than the vest.
The blouse is wetter than the vest.
The overcoat is softer than the blouse.
The shawl is dryer than the kilt.

(Manuscript received January 9, 2004;
revision accepted for publication May 23, 2005.)