

# Memory for mental models of spatial descriptions: An episodic-construction-trace hypothesis

STEPHEN J. PAYNE

University of Wales, Cardiff, Wales

Three experiments, based on Mani and Johnson-Laird's (1982) study, tested memory for spatial descriptions. In Phase 1 of each experiment, subjects judged whether diagrams matched verbal descriptions that were either determinate or indeterminate. In Phase 2, subjects attempted to recognize the descriptions studied in Phase 1. Mani and Johnson-Laird reported a crossover: Gist memory was better for determinate descriptions and verbatim memory was better for indeterminate descriptions. This crossover suggests that mental models are remembered. All three new experiments failed to replicate the crossover, challenging whether models are preserved in episodic memory. It was hypothesized instead that episodic memory records the mental operations by which models are constructed. This hypothesis accounted for the findings of all three experiments. In Experiments 1 and 2, performance varied with the construction-trace overlap between recognition test descriptions. In Experiment 3, performance was depressed if the order of sentences within descriptions was altered between study and test.

Johnson-Laird's (1983) conception of a mental model is an important construct in the psychology of language comprehension and reasoning, and may speak to broader issues such as the way people use machines. Yet the central idea is simple: A mental model is an *analog* representation of a real or imagined state of affairs. Analog representations, in contrast to propositional representations, do not have an arbitrary structure: The structure of an analog representation preserves the relational structure of the represented state of affairs.

Distinguishing analog from propositional representations empirically is controversial. Indeed, at one time, Anderson (1978) argued that it is in principle impossible (for critiques of his argument, see Johnson-Laird, 1983, and Pylyshyn, 1979). However, a great deal of recent empirical evidence supports the hypothesis that people construct and consult analog mental models: Consider, for example, the evidence on deductive reasoning presented by Johnson-Laird and Byrne (1990).

Furthermore, this empirical evidence is supported by computational and philosophical arguments. Anderson (1983) found it convenient to suppose imagistic representations in his ACT\* theory of memory representation. More recently, Newell (1990) has argued that mental models offer different computational advantages and disadvantages to propositional representations: Mental models are more limited in scope but easier to process,

and models do not suffer from the frame problem, because when the situation being modeled changes, it is obvious how an analog model should be updated. (This is famously not true for arbitrary propositional representations; see, e.g., Pylyshyn, 1987.) Finally, McGinn (1989) has argued that analog representations offer a better foundation for the notion of mental content than do propositional "languages of thought."

Currently, perhaps, the case for mental models is strongest for *spatial* mental models, for which converging evidence from different empirical paradigms can be marshaled. The well-known image scanning results of Kosslyn and colleagues (see, e.g., Kosslyn, 1980) offer support for regarding mental images as analog visuospatial representations, although their interpretation remains highly controversial. More recent results using a similar paradigm, and again interpreted in terms of analog spatial mental models, in this case derived from text, are reported by Denis and Cucode (1989).

Investigating representations of continuous dimensions is not the only way to empirically distinguish analog from propositional representations. Ehrlich and Johnson-Laird (1982) measured reading speed, listening speed, and memory for categorical spatial descriptions. Referentially continuous descriptions, in which adjacent sentences shared a referent, were easier to process according to all three measures, suggesting that subjects attempted to integrate propositions into a single coherent mental model. Morra (1989, Experiment 3) replicated the effect on reading times and memory with more closely controlled materials. Byrne and Johnson-Laird (1989) showed that the difficulty of deducing unstated relations from descriptions of spatial arrays was predicted by the number of spatial arrays that subjects had to consider—a result predicted by the theory of analog spatial mental models but not by

---

Helpful points, at various stages in the life of this research, have been made by Peter Bibby, Ruth Byrne, Alan Collins, Judi Ellis, Sue Gathercole, Phil Johnson-Laird, and Geoff Underwood. Special thanks to Christian Vaughan, whose undergraduate dissertation prompted this research. The experiments reported here were presented at the meeting of the Experimental Psychology Society in York, England, in July 1992. Correspondence should be addressed to the author at the School of Psychology, University of Wales, P.O. Box 901, Cardiff CF1 3YG U.K.

the logical form of the spatial descriptions. Finally, Franklin and Tversky (1990) and Franklin, Tversky, and Coon (1992) reported that subjects who read narratives that describe spatial scenes from the point-of-view of a single character form spatial mental models that incorporate that character's perspective.

### Are Mental Models Remembered?

If one accepts the utility of analog mental models in comprehension and reasoning, an important question about the psychological status of mental models remains open. Mental models are constructed, considered, and manipulated "to think with," but how are mental models remembered? Most of the wide ranging empirical work on text comprehension or deduction simply does not address this issue.

The case is clearest for deduction, since long-term memory is never tested. Experimental tests of model-based theories rely on the idea of analog mental models being constructed from premises, but subjects in these reasoning experiments do not need to remember their mental models once they have reported conclusions.

The same is true of many discourse comprehension experiments, such as those cited above of Denis and Cucode (1989) and of Franklin and colleagues (Franklin & Tversky, 1990; Franklin et al., 1992). Those studies that do implicate analog mental models do not, on my reading, address the relationship of mental models to episodic memory. Just as in experiments on deductive reasoning, the experimental procedures only require subjects to process spatial descriptions one at a time, uninterrupted, without any intervening alternative tasks that might compete for temporary working storage.

Consider, for example, the work of Franklin, Tversky, and colleagues (Franklin & Tversky, 1990; Franklin et al., 1992). In their experiments, subjects have to read rich descriptions of some physical scene from the point of view of an observer and then recall as quickly as possible the names of objects, cued by their positions relative to the observer. The basic finding is that response time can be predicted from the relative orientation of the observer in the described scene. These findings seem to offer strong evidence for spatial mental models (and, further, to identify them as "viewer centered"), but they do not implicate long-term episodic memory. It is true that subjects must repeatedly consult their model of each spatial scene for several minutes, but there is no reason to suppose that subjects do not *maintain* (or "refresh," somehow) their mental model, in working memory, throughout this period. Once subjects abandon one mental model for another—that is, once they move on to the next part of the experiment and begin to read some completely novel description—they are never called upon to remember the earlier model.

In other work on mental models in text comprehension, the key hypothesis is that comprehension goes beyond the text-level propositions to some understanding of the situation that is described. Although such tests will often

implicate long-term memory, it is the content of memory that is at stake, not its representational form. A model of the described situation can be distinguished from the propositions in the text, without concern for whether the situation is represented in analog form. Indeed, van Dijk and Kintsch's (1983) formulation of a "situation model," which is similar to Johnson-Laird's (1983) theory in its stress on the use of inferential processes and world knowledge, does not make claims about an analog representational format. As van Dijk and Kintsch put it, the fundamental insight is that "to understand the text we have to represent what it is about" (p. 337).

In support of this insight, several studies have shown that spatial descriptions of various kinds can lead to long-term representations of the spatial situation that is described (see, e.g., Glenberg, Meyer, & Lindem, 1987; Morrow, Bower, & Greenspan, 1989; Morrow, Greenspan, & Bower, 1987). To explain results like these, one needs to suppose that readers infer unstated properties of the situation described in the text and that they remember these properties; one need not suppose any analog mental model.

This argument can be illustrated by considering the classic experiments of Bransford and colleagues (Bransford, Barclay, & Franks, 1972; Bransford & Franks, 1971). Bransford et al. show that sentence memory cannot adequately be explained purely in terms of the linguistic meaning of sentences but is instead a function of situations that the sentences describe. In one experiment, subjects were likely to make false-positive recognition responses to a new sentence, if this new sentence used a different pronoun, but described a spatial situation identical to that in some remembered sentence. For example, subjects who had heard, "Three turtles rested on a floating log, and a fish swam beneath them" tended to falsely think they had heard, "Three turtles rested on a floating log, and a fish swam beneath it." However, subjects who had heard, "Three turtles rested beside a floating log, and a fish swam beneath them" tended not to think they had heard, "Three turtles rested beside a floating log, and a fish swam beneath it."

This finding can be explained by assuming that subjects remember an analog mental model of the situation described by the sentences that they hear. However, as Bransford et al. (1972) point out, in discussing mental imagery, the results do not dictate what form the mental representation of the situation must take, only that the representation contains information that is not explicitly present in the linguistic inputs.

What might the relationship be between analog working-memory models and long-term memory? There are two broad alternatives. Perhaps the simplest assumption is that mental models exist as a special representational form in both working and long-term memory. The second major possibility is that mental models are redescribed propositionally in long-term memory.

This second idea, recoding, is favored by Glenberg and Langston (1992), who are among the few to recognize

the issue. They report experiments investigating the effects of a picture of a four-step procedure on the long-term memory for the procedure. They argue that pictures help the formation of a mental model in working memory. However, they prefer to limit the special representational form to working memory, positing some propositional trace of such a model in long-term memory. They regard such a proposal as more parsimonious, since it can "take advantage of the tremendous literature supporting propositional representational formats." However, they offer few ideas about how such a propositional trace might be constructed. Furthermore, their empirical evidence could equally well be explained by assuming that longer term recall does indeed rely on a mental model representation, if one is available.

Johnson-Laird (1983) apparently believes that the structure-preserving analog quality of mental models is a property of enduring knowledge representations as well as of working-memory models. Johnson-Laird uses a single experiment to bear the weight of his central thesis that there is an empirical distinction between mental models and propositional representations. The critical experiment is the second reported by Mani and Johnson-Laird (1982) in their investigation of the mental representation of spatial descriptions. The interpretation of this experiment relies on the construct of analog mental models in long-term episodic memory.

Mani and Johnson-Laird (1982) investigated memory for spatial descriptions, such as "The knife is to the right of the fork." For such descriptions, clear differences exist between the expressive power of propositional and analog representations. For example, an analog representation is incapable of expressing "vague" spatial propositions like "the fork is next to the spoon" (Pylyshyn, 1973): We can say that such propositions are *indeterminate* with respect to a model, because they are true of more than one distinct spatial configuration. The Mani and Johnson-Laird experiments exploited this critical distinction between analog and propositional representations. Their procedure will be described in some detail, since the detail is critical for the arguments and new experiments to follow.

In the first phase of the experiments, subjects were presented with a series of descriptions, each comprising multiple sentences describing a configuration of multiple objects. (In the second study, which is the one on which we focus, four sentences described the layout of five objects.) For example, subjects might read the following sentences:

The bookshelf is to the left of the chair.  
The table is to the right of the chair.  
The sideboard is in front of the bookshelf.  
The bed is in front of the chair.

After the subjects read each description, the description was removed when they were ready, and the subjects were shown a diagram, illustrating some arrangement of the same object names. The subjects had to state whether or not the description was true of the arrangement (it was

true on half the trials). The critical manipulation in this first phase of the experiment was the determinacy of the description. The sentences above are determinate (ignoring scale and assuming symmetrical placements) because they are true of just one configuration. But if a simple change is made to just one sentence by substituting a single object-name, the description becomes indeterminate because the relative position of two objects is uncertain, so that more than one mental model can be constructed. For example,

The bookshelf is to the left of the chair.  
The table is to the right of the bookshelf.  
The sideboard is in front of the bookshelf.  
The bed is in front of the chair.

After practice tasks, each subject viewed and judged four determinate and four indeterminate descriptions. Each description described the layout of five objects in a rotation of an F-shape (half of each were followed by correctly matched diagrams; half were followed by incorrectly matched diagrams depicting the same objects in a different F-shape layout).

In a second phase of the experiment, the subjects were given a surprise recognition memory test. For each item of the test, the subjects were shown a previous description accompanied by three foils. They had to rank these four descriptions according to how closely the descriptions resembled a description they had seen in the first phase. The foils were constructed according to a fixed recipe: Each foil contained four sentences of the same general form as the four sentences in the original description. One foil, the "inferrable" description, contained different sentences from the original but described the same configuration of objects; the other two "confusion descriptions" described a different F-shape layout of the same objects.

The subjects' recognition performance was scored as exhibiting verbatim memory if the original description was ranked ahead of the inferrable description and gist memory if the original and the inferrable descriptions, in either order, were ranked ahead of the two confusion items. Note that under these operationalizations, a single ranking can exhibit both verbatim and gist memory. The key result was a crossover. Gist memory was better for determinate descriptions (average of 88%) than for indeterminate descriptions (average of 58%). Verbatim memory was better for indeterminate descriptions (average of 88%) than for determinate descriptions (average of 68%).

Mani and Johnson-Laird (1982) interpret this finding as follows. During Phase 1, readers of the descriptions attempt to form mental models—that is, analog representations of the described configurations. But when the description is indeterminate, this approach is problematic. (The subsequent reasoning experiment by Byrne & Johnson-Laird, 1989, lends support to this conjecture.) Consequently, on these occasions, subjects need to hold in mind the original verbal propositions in order to do the diagram-matching task. To perform the recognition task, therefore, subjects are likely to rely on a remem-

bered mental model for the determinate descriptions but on remembered propositions for the indeterminate descriptions. Models will be better remembered than will propositions (Mani & Johnson-Laird simply recruit depth of processing to support this point), but they do not allow the inferrable description to be discriminated from the original. Hence, for determinate descriptions, gist memory is boosted, but verbatim memory is depressed. When the memory is propositional, the original and inferrable descriptions can be discriminated, thus verbatim memory is boosted for indeterminate descriptions.

This argument certainly offers a possible explanation for the crossover result. However, Johnson-Laird (1983) argues further that "this 'cross-over' effect is impossible to explain without postulating at least two sorts of mental representation" (p. 162). Yet the crossover *can* be explained by a unitary representation with the right properties. To understand how this is possible, consider how performance on the recognition tests varies with the "psychological distance" between the inferrable and original descriptions (i.e., the similarity between the descriptions as they are represented in the subject's mind). As the psychological distance increases, verbatim memory will be boosted, because it depends on the subject being able to rank the original description ahead of the inferrable description. At the same time, gist memory will be depressed, because it depends on the subject being able to distinguish both the original and inferrable descriptions from the two confusion descriptions. Obviously, as psychological distance between original and inferrable descriptions decreases, the opposite effect on recognition performance will be observed: Verbatim memory will be depressed, and gist memory will be boosted. In these terms, Johnson-Laird's explanation of the determinacy-crossover effect is that, for a mental model representation, psychological distance between original and inferrable is zero, whereas for a propositional representation, psychological distance between original and inferrable is nonzero. However, there is no a priori reason why some unitary representation should not allow psychological distance between original and inferrable to vary. Such a representation may, therefore, explain the crossover effect, just so long as the psychological distance between inferrable and original descriptions is on average greater for the indeterminate than for the determinate descriptions.

In the course of this article, a unitary representation will be proposed in which the psychological distance between original and inferrable descriptions can vary substantially, despite the constant recipe for constructing inferrable descriptions. This allows variation in gist and verbatim memory to arise through an artifact of the materials. Indeed, in Experiment 1, reported below, exactly such an artifactual result is reported. Experiment 1 is an attempt to replicate the Mani and Johnson-Laird (1982) result while adopting somewhat stricter recipes for the construction of recognition-test descriptions in order to remove the possibility of artifactual variations in recognition performance.

## EXPERIMENT 1

### Method

**Materials.** Descriptions and diagrams for Phase 1 followed the Mani and Johnson-Laird (1982) recipes. Every description comprised four sentences, each of which described the spatial relation (in front, behind, to the right, or to the left) between two objects, and which together described the arrangement of five objects in one of the rotations of an F-shape. Each sentence, except for the first, introduced one "new" object, thus preserving referential continuity. Half of the determinate and indeterminate descriptions were followed by a matching diagram (henceforth, these descriptions will be referred to as "diagram-matching descriptions"); the other half were followed by a diagram showing the same objects in a different rotation of an F-shape (henceforth, "non-diagram-matching" descriptions).

Materials for the second phase were constructed as follows. There were eight separate test items, one for each of the descriptions judged in Phase 1. Each item comprised an original description, an inferrable description, and two confusion descriptions. Each description comprised four sentences describing the layout of the same five objects. The inferrable descriptions were constructed according to the Mani and Johnson-Laird (1982) recipe, with one minor change. As before, there was one new sentence, true of the arrangement originally described, but not used in the original description. Only one sentence was carried over unchanged from the original description (instead of two); two sentences appeared in converse form (instead of one). This change decreases the perceptual similarity between original and inferrable descriptions, reducing the potential for very superficial recognition strategies. It also reduces the linguistic similarity between original and inferrable, so it might be expected to boost verbatim scores and, perhaps, depress gist scores. But it has no influence on the hypothesized relationship between determinacy and recognition performance. The single unchanged sentence was first, second, third, and fourth in one each of the four determinate items, and likewise for the indeterminate items.

Another change was made to facilitate a stricter recipe for constructing the two confusion descriptions. The first confusion description was created from the original description by simply interchanging three of the object names while keeping two the same. This satisfies part of Mani and Johnson-Laird's (1982) specification because the description contains the same relational terms in the same order as the original but describes a different layout of objects. However, the shape of the described configuration (i.e., the rotation of an F-shape) is preserved, whereas it seems that Mani and Johnson-Laird used a different shape for both of their confusion descriptions (they did in Experiment 1 and constructed them along "the same general lines" in Experiment 2).

The new construction recipe allows, by choosing appropriate object interchanges, the single common sentence from the original and inferrable descriptions to be maintained in the two confusion descriptions, which was done consistently throughout. The second confusion description applied the same object-interchange transformation to the inferrable description. Again, this complies with the Mani and Johnson-Laird (1982) recipe because it describes a different layout and makes use of relational terms different from those in the original. But the recipe additionally ensures that the two confusion descriptions always describe the same configuration of objects. The other important feature of our materials is that a single sentence is common to all four descriptions in every recognition-test item and is the only shared sentence. Both of these changes work against the plausible recognition-test strategy of preferring descriptions that are more superficially similar to each other, independent of any memory for descriptions seen in Stage 1.

All of the descriptions preserved the property of referential continuity, introducing at most one new object per sentence after the first sentence. The descriptions for determinate and indeterminate

cases were of the same form except for the changes resulting from the single change of noun phrase in the original description. All recognition-test descriptions presented with determinate originals were determinate; those presented with indeterminate originals were indeterminate.

Appendix A lists the complete set of recognition-test descriptions for one determinate and one indeterminate case.

**Procedure.** The experimental procedure, including instructions to subjects, was modeled precisely on Mani and Johnson-Laird's (1982) report, except that it was run on an interactive computer program that presented all materials and recorded subjects' responses. In Phase 1, all subjects were given two practice items (one a good match, the other a bad match) and then were told to judge the eight test items in random order. On each item, the subjects clicked a mouse to indicate that they were ready to move from description to diagram and then selected "Good Match" or "Bad Match," which appeared as "buttons" on the screen, with the mouse. The subjects were given no feedback on their diagram-matching responses. Immediately after completing Phase 1, the computer displayed instructions for Phase 2. The eight recognition items of Phase 2 were presented in random order, and, within each item, the four descriptions (original, inferrable, and the two confusion descriptions) were ordered randomly in a single column in the center of the display. To fit all four descriptions on a single display, the font used in Phase 2 was smaller than the font used in Phase 1. For each item of Phase 2, the subjects ranked the four descriptions by pointing to each with the mouse. During a ranking, the subjects could change their minds and start the ranking again by selecting a "Cancel" button.

**Subjects.** Fifteen first-year undergraduates at Lancaster University were each paid £2 for serving as subjects.

**Scoring of data, and predictions.** To assess performance in the recognition phase, Mani and Johnson-Laird (1982) report the  $P$  statistic, which is a measure of the correlation between the subjects' rank ordering of descriptions and the ideal ordering. However, it is unclear what the better ordering of the two confusion descriptions is. Furthermore, the critical argument from data to the theory of mental models concerns the separation of verbatim and gist memory performance. Mani and Johnson-Laird considered responses as exhibiting verbatim memory whenever the original was ranked ahead of the inferrable, but this scheme is hard to justify—if one or both confusion descriptions are ranked first, it seems strange to credit subjects with any memory success at all. Consequently, we report verbatim memory only on those trials for which the original description was ranked first. Note that the predictions of the remembered-mental-models account are unchanged by this amendment to scoring procedures. (Furthermore, all of the analyses reported below, in all experiments, were repeated with the Mani and Johnson-Laird scoring rubric. The pattern of findings is unaffected by this manipulation, as one would expect if Mani and Johnson-Laird's reported verbatim scores have simply been elevated by chance performance.)

Following Mani and Johnson-Laird (1982), the subjects' responses were scored as exhibiting gist memory whenever the original and inferrable, in either order, were ranked ahead of the two confusion items.

To repeat the predictions from the Mani and Johnson-Laird (1982) account: Memory for determinate descriptions will rely on a mental model, whereas memory for indeterminate descriptions will rely on the verbal propositions. Consequently, verbatim performance will be better for indeterminate descriptions than for determinate descriptions, but gist performance will be better for determinate descriptions than for indeterminate descriptions.

## Results

In Phase 1, erroneous judgments about the diagrams accounted for 20.8% of the responses. The errors did not

differ significantly as a function of the determinacy of the descriptions or the goodness-of-match of the diagrams. The subjects received no feedback on the results of their diagram judgments, so there is no reason to suppose that correctness of judgment will influence Phase 2 performance. Nevertheless all of the analyses reported below, for all three experiments, were repeated after excluding those items for which the subjects judged the goodness-of-match of the diagram incorrectly. The patterns of significance were completely unaffected, hence these analyses are not reported.

In the Phase 2 recognition test, there was an absence of any effect for the determinacy of the description. Table 1 shows the mean gist and verbatim scores, expressed as percentages, separated according to determinacy of descriptions and match of the diagram presented in Phase 1. Each cell in the table represents the average of two responses per subject.

Differences of verbatim and gist scores between determinate and indeterminate descriptions and between diagram-matching and non-diagram-matching descriptions were tested with four separate Wilcoxon's tests, pooling across the two values of the other variable in each case. The only significant difference was between diagram-matching and non-diagram-matching descriptions for gist recognition scores. Surprisingly, the subjects' gist memory performance was significantly better for non-diagram-matching descriptions (Wilcoxon's  $T = 9$ ,  $n = 12$ ,  $p < .02$ ). Verbatim performance was better for diagram-matching descriptions, but this difference was not reliable.

No tests of interaction were performed, but the shape of the interaction between verbatim and gist performance and determinate versus indeterminate descriptions was, in any case, in the direction opposite to that of the crossover interaction reported by Mani and Johnson-Laird (1982). Verbatim scores were (nonsignificantly) better for determinate descriptions, whereas gist scores were (nonsignificantly) better for indeterminate descriptions.

## Discussion

Why might it be that the crossover effect of determinacy observed by Mani and Johnson-Laird (1982) is not apparent in this experiment? One possibility is that, by

**Table 1**  
Results of Experiment 1

Diagram	Gist	Verbatim
Determinate		
Matching	43	90
Nonmatching	50	67
Indeterminate		
Matching	37	77
Nonmatching	70	67

Note—The above data show the mean percentage of trials on which the subjects scored gist and verbatim recognition performance according to whether the description studied in Phase 1 was determinate or indeterminate and whether it was presented before a matching or a nonmatching diagram. Each cell averaged over two responses per subject.

more strictly controlling the Phase 2 materials, some spurious artifact has been eliminated. However, there is a much less damaging possibility, which might partly explain the different findings in the two experiments. One substantive difference between the new items and Mani and Johnson-Laird's items is the "shape" of the layout described in the two confusion descriptions. In Mani and Johnson-Laird's experiment, each of the two confusion descriptions described an F-shape different from that of the original (and inferrable) description (it is not clear whether both confusion descriptions described the same shape). In Experiment 1, all four descriptions described the same shape.

If recognition performance is based on a partially remembered mental model, perhaps the *shape* of the mental model, rather than the specific contents (i.e., positions of individual objects), is what is preserved in memory and thus is what is used to make judgments. In Mani and Johnson-Laird's (1982) experiment, memory for shape would have led to gist recognition. In Experiment 1, memory for shape only will not support any recognition, since the shape of all four recognition-test items is the same. If this account is broadly correct, we might expect reduced gist scores in the new experiment, relative to those reported by Mani and Johnson-Laird, at least for the determinate items, for which remembered mental models have been proposed to underlie gist recognition performance. This was indeed what we observed. The verbatim scores in Table 1 are very close to those reported by Mani and Johnson-Laird, despite the stricter scoring employed in the current study. The gist scores for indeterminate items are also close to those reported by Mani and Johnson-Laird, but the gist scores for determinate items are quite substantially depressed.

The second aspect of the results that begs some explanation is the significant difference on gist memory performance between diagram-matching and non-diagram-matching descriptions. A priori, one might imagine that diagram-matching descriptions would show some advantage because, for these descriptions only, memory for the diagram itself would support gist memory. The significant difference in the opposite direction is hard to explain. However, the new hypothesis about memory for mental models that is presented below shows how the effect could be caused by an experimental artifact.

### A Construction Trace Account of Memory for Mental Models

Despite the failure to replicate Mani and Johnson-Laird's (1982) results, one might argue that the depressed gist performance that we have observed, in the face of recognition-test items that all preserve the shape of the layout described in the to-be-remembered descriptions, is broadly consistent with an account based on remembered mental models. However, discussions with subjects immediately after the experiment, together with introspections on performing the experimental task, suggested a novel hypothesis about what is remembered from Phase 1

of the experiment. Suppose that people do indeed attempt to construct mental models in order to perform the diagram-matching task, but that neither these mental models nor the verbal propositions from which they are constructed are remembered. Instead, subjects remember the mental operations that they must perform in order to construct the model. I call this remembered list of operations the *episodic construction trace* of the model.

Johnson-Laird (1983) describes a simple computer program that constructs analog mental models (represented as two-dimensional arrays) from verbal propositions. This program was reimplemented to lend some rigor to the idea of an episodic construction trace. The new program builds arrays according to the Johnson-Laird procedure but discards the array once it has been built and used for reasoning (judging diagrams). During the building of the array, lists of symbols that describe the array-building operations are constructed and stored. To the extent that Johnson-Laird's procedure for constructing mental models is general, and I believe it to be very general, so is the proposed construction-trace model of memory. However, to clearly describe the construction trace, it is necessary to relate it to the verbal materials currently under consideration. The construction trace for the current spatial descriptions preserves the following distinctions: (1) It distinguishes the starting pair of objects that are first placed in the model array; (2) for all subsequent propositions, it distinguishes the new object from the object that is already in the array (the syntactic convention is simply to place the new object at the head of the list), and it records the relation of new to old; (3) if the old object is already placed in the same relation to another object in the array, it records the relation of the new object to both already-situated objects; and (4) it records any clashes, in which the relative position of two objects is indeterminate, as a modifier of the relation that first produces the clash. Appendix B lists an example of the episodic construction trace for two of the original descriptions used in Experiment 1.

Each assumption of the construction-trace model, as it relates to the current verbal materials, can be given an informal justification with respect to Johnson-Laird's (1983) construction algorithm, which itself is extremely general. Constructing a mental model, according to this algorithm, is an incremental process in which new elements are placed in an array sequentially as they are encountered in the text. The first two elements have a special status because they signal the beginning of model construction, they are mentioned in the same proposition, and only their relation to each other is initially known. For these reasons, it is plausible that the placing of these two objects in the model is identified in the memory of the construction process (Assumption 1). Subsequent propositions all require that a single new object is added to the existing model. What is important to the model-building procedure, therefore, is not necessarily the relation that is expressed in the current verbal proposition but rather the relation of the new object to the old ob-

ject(s) that is expressed in that proposition. Consequently, the construction trace will sometimes record the spatial relation opposite to that in the proposition (Assumption 2). When the current model contains two objects that are in some relation, and the new proposition requires that a new object be placed in that relation to one of those objects (e.g., when a new object must be placed to the right of an object that has a right/left relation with another object in the model), the construction process must determine the relation of the new object to the other element of the established pair that is not mentioned in the proposition but is already situated in the model. Consequently, the construction trace records the position of the new object relative to both objects in the pair (Assumption 3). For some propositions, in indeterminate descriptions, it is not possible to place the new element relative to the unmentioned element of the pair, since this is what makes the descriptions indeterminate. Consequently, the construction trace records a clash (Assumption 4).

In general, then, the episodic construction trace does not rely on any ad hoc assumptions over and above Johnson-Laird's (1983) general analysis of incremental mental model building. Rather, the construction trace records only those mental operations that are necessary to construct a mental model.

The construction-trace hypothesis limits the analog representational form of mental models to temporary working memory. The construction trace in long-term episodic memory is propositional in form, in the sense that an arbitrarily structured list of symbols is used to represent knowledge. Nevertheless, as will become clear, the construction trace is quite distinct from the verbal propositions from which the mental model was constructed.

Suppose, then, that Phase 2 recognition performance in Experiment 1 (and in Mani & Johnson-Laird's, 1982, experiment) is based on the episodic construction trace. To perform the recognition task, subjects must construct new working-memory mental models for each of the four descriptions and compute the "trace overlap" between the new and remembered construction traces. Assuming a crude numeric index of overlap based on the number of common sublists, if memory for Phase 1 is perfect, the trace overlap with the original description will be 4, but any loss of information will reduce this to 3 or less. Trace overlap between the original and inferrable descriptions can vary between 3 and 0, despite the uniform recipe for construction of inferrable descriptions. Appendix B illustrates this variance. Trace overlap between the original descriptions and the confusion items in Experiment 1 is either 0 or 1 because of the single shared sentence.

It must be noted that the trace overlap between two descriptions is quite distinct from the similarity of the descriptions with respect to common propositions. The examples in Appendix B illustrate this point. The propositional similarity between each original description and its corresponding inferrable description is constant, being determined by the experimental recipe for inferrable descriptions. (In the examples of Appendix B, the number of common propositions is two in every case.)

In the construction-trace model, partial forgetting can be simulated by random probabilistic deletion of sublists from the stored construction trace. By adjusting the likelihood of such deletions, it is easy to simulate the observed levels of performance. The computer simulation ranks the four recognition-test items in order of decreasing trace overlap. Whatever these overall levels, verbatim recognition will improve and gist recognition will diminish as the trace overlap between the original and inferrable descriptions decreases.

The numeric trace overlap offers an explanation for the curious significant difference between descriptions that were presented with matching versus nonmatching diagrams. Inspection of the experimental materials shows that, by accident, all descriptions that were presented with nonmatching diagrams had a trace overlap of 3 with their corresponding inferrable descriptions, whereas all matching descriptions had a trace overlap of 1.

The episodic-construction-trace hypothesis can also account for the main findings of Bransford et al. (1972), described above. Appendix C shows the construction trace that would result from initially forming a mental model of the two versions of the turtles-log-fish sentence. If, during the recognition task, subjects used trace overlap to guide their response, false positives would increase as trace overlap between new and old sentences increases.

## EXPERIMENT 2

Experiment 2 had two main goals. The first goal was to test whether the Mani and Johnson-Laird (1982) crossover would reappear if the confusion items in the recognition test described a shape different from that of the original, as well as a different layout of objects. The second goal was to directly test the construction-trace hypothesis by deconfounding trace overlap from the goodness-of-match of the diagram.

### Method

**Materials.** The construction of materials for the recognition phase followed exactly the Mani and Johnson-Laird (1982) recipes. The inferrable description used two same sentences, one converse sentence, and one novel sentence to describe the same layout as the original. Both confusion items described an F-shape different from those in the original and inferrable descriptions; the first confusion item used the same relation terms in the same order as in the original description. The trace overlap between original and inferrable descriptions was either 1 or 3, and this variation was counter-balanced across determinate versus indeterminate and diagram-matching versus non-diagram-matching descriptions.

**Procedure.** The procedure was identical to that of Experiment 1.

**Subjects.** Nineteen undergraduates at the University of Wales, Cardiff, were each paid £3 for serving as subjects.

### Results

The data were treated in the same way as in Experiment 1.

In Phase 1, errors in classifying diagrams constituted 21.0% of the responses. The errors did not vary significantly as a function of the nature of the descriptions or the diagrams.

**Table 2**  
**Results of Experiment 2**

Diagram	Trace Overlap	Gist	Verbatim
Determinate			
Matching	1	53	79
	3	74	79
Nonmatching	1	21	68
	3	47	63
Indeterminate			
Matching	1	26	63
	3	74	58
Nonmatching	1	37	79
	3	53	58

Note—The above data show the mean percentage of trials on which the subjects scored gist and verbatim recognition performance according to whether the description studied in Phase 1 was determinate or indeterminate, whether it was presented before a matching or a nonmatching diagram, and whether the trace overlap between the original description and the “inferred” recognition-test item is 1 or 3. Each cell averaged over one response per subject.

Table 2 shows the mean verbatim and gist recognition memory scores, expressed as percentages, according to the determinacy, diagram match, and original/inferred trace overlap of the items. Each cell in the table represents the average of one response per subject. Significance tests were carried out separately for verbatim and gist scores, for the main effects of determinacy, diagram match, and trace overlap, using six Wilcoxon *T* tests.

There were no significant differences between determinate and indeterminate descriptions. Nor were there any reliable differences between diagram-matching and non-diagram-matching descriptions.

Once again, the shape of the interaction between verbatim and gist performance and determinate versus indeterminate descriptions was in the direction opposite to that of the crossover interaction reported by Mani and Johnson-Laird (1982). Verbatim scores were (nonsignificantly) better for determinate descriptions, as were gist scores, but the difference was greater for verbatim scores.

Gist memory was significantly greater for those items for which the trace overlap between original and inferred descriptions is 3 than it is for those items for which the original/inferred trace overlap is 1 ( $T = 5, n = 14, p < .01$ ). Verbatim memory was better for those items for which the original/inferred trace overlap is 1, though not significantly so.

### Discussion

As in Experiment 1, the observed overall levels of gist memory are lower than those reported by Mani and Johnson-Laird (1982). However, the difference in this case cannot be attributed to the shape of the layout described by confusion descriptions. Despite being depressed, levels of gist performance in all conditions are well above chance (chance performance is 1/6).

Once again, there is no hint of the Mani and Johnson-Laird (1982) crossover result: The determinacy manipu-

lation makes no appreciable differences in either verbatim or gist recognition performance. There is, therefore, no evidence that any mental models constructed in Phase 1 are remembered during Phase 2.

The curious effect of diagram match, observed in Experiment 1, has disappeared now that it has been separated from trace overlap. The strong remaining effect for trace overlap on gist recognition is consistent with the episodic-construction-trace hypothesis. Indeed, because trace overlap is independent of any other available measure of similarity between original and inferred descriptions (such as the meaning of the descriptions or the number of common propositions, both of which are constant throughout the experiment), it is hard to imagine how to explain the reliable effect of the trace overlap manipulation (and its accidental manipulation in Experiment 1) without invoking some version of the construction-trace hypothesis.

Together with the results of Experiment 1, these findings suggest that the large significant differences between determinate and indeterminate descriptions observed by Mani and Johnson-Laird (1982) may have been an artifact, perhaps arising in exactly the same way as the differences between matching and nonmatching descriptions in Experiment 1—namely, accidental variations in the trace overlap between original and inferred descriptions. Some additional support for this conjecture is provided by the observation that it is possible to produce indeterminate inferred descriptions that have a trace overlap of 0 with the original, whereas for determinate descriptions, the minimum original/inferred trace overlap is 1. Appendix B shows an example of zero trace overlap despite the carryover of two unchanged sentences.

### EXPERIMENT 3

The episodic construction trace is highly sensitive to the order in which propositions are considered during the construction of a mental model. Although the construction-trace list can itself be treated as unordered, the symbols within each sublist explicitly encode order relations by marking which objects start the model construction and which object of any pair is the new addition to the model.

A mental model itself is, of course, completely insensitive to the order of the process by which it is constructed. If the sentences (and thus the propositions) in any of the descriptions are reordered, the associated mental model is unchanged.

A remembered list of verbal propositions might or might not be ordered, but it surely can support recognition performance that is order-insensitive. One who remembers a list of propositions can consider whether it contains the same propositions as does a novel description independently of the order within the list.

These considerations lead to a new way of testing the construction-trace hypothesis. If the original descriptions that are presented in Phase 2 are reordered, the trace over-



lap between the remembered construction trace and the "original" description will be reduced. Appendix D shows the construction traces for an original and a reordered description. Therefore, if recognition performance is based on the construction trace, it will be damaged. However, if recognition performance is based on a remembered mental model, or on a remembered list of propositions, the reordering should have no effect, provided that subjects are instructed to ignore sentence order while making their judgments.

### Method

**Materials.** The materials were the same as were used in Experiment 1, except that half of the original descriptions (two determinate and two indeterminate, one of each with matching and one with nonmatching diagrams) were re-presented in Phase 2 with the four sentences reordered. The reordering was always done so as to maintain referential continuity.

**Procedure.** The only alteration to the procedure of Experiments 1 and 2 concerned the instructions given to the subjects at the beginning of Phase 2. The subjects were explicitly instructed to ignore the order of sentences within each description while performing the recognition/ranking task.

**Subjects.** Fifteen undergraduates at Lancaster University were each paid £2 for serving as subjects.

### Results

In the diagram-judging phase, errors were made on 29% of the responses. There is no explanation for why this error rate should be higher than that in Experiments 1 and 2 or in Mani and Johnson-Laird's (1982) experiments. Errors were evenly distributed among determinate and indeterminate descriptions. There were more errors made in judging matching diagrams ( $T = 2, n = 7, p < .05$ ), suggesting a bias to respond "Bad Match."

Table 3 shows the mean verbatim and gist memory scores, expressed as percentages, separating determinate from indeterminate and stable from reordered items. Each cell in the table represents the average of two responses per subject.

Once again, significance tests of main effects were carried out using separate Wilcoxon's  $T$  tests. There was no effect of determinacy, on either verbatim or

gist memory performance. There was a significant effect of reordering for both verbatim ( $T = 15, n = 13, p < .025$ , one-tailed) and gist performance (not a single subject went against the trend of scoring worse gist recognition performance on reordered items,  $p = .5^{15}$ ).

For the first time in this series of experiments, the shape of the interaction between verbatim and gist performance and determinate versus indeterminate descriptions was in the direction of the crossover interaction reported by Mani and Johnson-Laird (1982). Verbatim scores were (nonsignificantly, but slightly) better for indeterminate descriptions, and gist scores were (nonsignificantly, but slightly) better for determinate descriptions. It is impossible to test this "interaction" in conventional ways (even with a nonparametric ANOVA), since gist and verbatim recall are not independent levels but, rather, different codings of the same performance data. The absence of any effect for determinacy on either gist or verbatim recall itself signals that the interaction is unlikely to be reliable. A further test was performed by counting the number of subjects who individually exhibited recall performance in the direction of the Mani and Johnson-Laird interaction—that is, subjects for whom the real value of the determinate-indeterminate difference was greater for gist than for verbatim recall. In fact, the responses of only 6 subjects showed this pattern, with 6 showing the opposite pattern (i.e., an "interaction" in the opposite direction), and 3 showing no difference. Once again then, for the third experiment in succession, there was no hint of a replication of the Mani and Johnson-Laird finding.

### Discussion

The depressed performance for items in which the original description is reordered cannot be explained if recognition relies on either a remembered mental model or a remembered list of propositions, because both these representations are unaffected by the reordering of sentences within descriptions. The construction-trace hypothesis can explain the effect, since a construction trace necessarily encodes the order in which propositions are considered, so that trace overlap is substantially reduced when descriptions are reordered.

Can any other model of memory explain the reordering effect? One possibility is that subjects remember some superficial aspects of the original materials that happen to be order-sensitive—for example, some syntactic or phonological encoding. If performance utilizes such a trace, then of course the reordering would be expected to have an effect. However, although such superficial traces could clearly underlie verbatim recognition performance, it is far from clear that any superficial encoding could be responsible for the levels of gist performance that are achieved in the stable-order condition. When a subject's ranking of items puts the original first and the inferrable second, so that both gist and verbatim recognition are observed on the same test item, the gist performance could conceivably be parasitic on verbatim memory and hence derive from a superficial encoding. However,

Table 3  
Results of Experiment 3

Reordered	Gist	Verbatim
Determinate		
Stable	57	70
Reordered	17	43
Indeterminate		
Stable	50	77
Reordered	13	37

Note—The above data show the mean percentage of trials on which the subjects scored gist and verbatim recognition performance according to whether the description studied in Phase 1 was determinate or indeterminate and whether or not the sentences in the original description were reordered in the recognition test. Each cell averaged over two responses per subject.

when gist memory is observed in the absence of verbatim memory (i.e., when subjects rank the inferrable description first and the original description second), it is hard to see how it could have arisen through just a superficial encoding. Yet such gist-without-verbatim-memory performance is reliably above chance in the stable-order condition (occurring on 7 out of the possible 15 responses that did not show verbatim memory; chance = 2/18;  $p < .01$ , binomial test) and is reduced to chance levels by reordering (occurring on 3 out of 33 responses). The construction-trace hypothesis can explain *both* of these phenomena.

One other possible explanation for the data should be considered—namely, that subjects construct and maintain multiple representations of each text: a mental model of the situation described, *and* a propositional representation close to the language of the text, *and* some superficial representation that is sensitive to word and/or sentence order. If recognition judgments are determined by some combination of the different representations (e.g., summing the activation across representations that is cued by the various recognition-test items), one might begin to simultaneously address good gist memory (supported by activation of the mental model) and the reordering effect (due to less activation of the superficial trace).

However, despite its initial promise, careful consideration reveals that such a multiple-representation model cannot address these particular data. The difficulty is that recognition performance on all tests is determined by forced-choice ranking of four items, and only the original item has been manipulated in the stable versus reordered conditions. If subjects were being asked to recognize the original item in isolation, multiple representations would indeed predict depressed performance under the reordering manipulation, for the simple reasons suggested above. But in the forced-ranking paradigm, I will argue that multiple representations (excluding something akin to the construction trace) cannot predict the observed data. To express these arguments, I assume that recognition-ranking responses are determined by the total activation of representations cued by recognition-test items. However, this is mainly a terminological convenience—I believe the argument is general to any account that relies on a simple coordination of supposed multiple representations to perform recognition.

Again, one argument hinges on consideration of those recognition tests on which a subject ranks the inferrable description first and the original description second, thus scoring gist but not verbatim recognition. In the stable-order condition, as noted above, these responses occur more often than one would expect by chance. Activation of a remembered mental model is the only way for multiple representations to explain such gist performance; a superficial, order-dependent trace can be doing no work on these responses. Consequently, a manipulation that affects use of the superficial trace but has no effect on use of a mental model cannot predict depressed gist-without-verbatim performance.

The multiple-representation framework can predict lower total activation of memory cued by the reordered original description, but it *cannot* predict lower relative activation from the reordered original than it can from the inferrable description (as they share the same mental model), and therefore it cannot readily predict a large depression of gist scores due to reordering of the original description. The construction-trace hypothesis, however, with its assumption of probabilistic maintenance of each element in the trace, readily predicts lower activation of the trace from the reordered original than from the inferrable description and therefore can predict the observed effect of reordering on gist memory scores.

Together, these arguments weigh heavily against a multiple-representations (models plus propositions) explanation of the data of Experiment 3. It may still be possible to construct such an explanation, but I have not been able to do so. (Furthermore, to compete with the construction-trace hypothesis, any such explanation should also address the effect of trace overlap in Experiments 1 and 2.)

## GENERAL DISCUSSION

The article by Mani and Johnson-Laird (1982) is unusual in the experimental literature on mental models because it suggests that mental models are a special representational form, not just in working memory but in longer term episodic memory as well. However, the three experiments reported above, including one that attempts to be a precise replication (Experiment 2), fail to replicate the differences in the processing of determinate and indeterminate spatial descriptions that provoked Mani and Johnson-Laird's claims about episodic mental models.

One might be inclined to question the empirical status of analog representations in general, but as noted in the introduction, there is by now plentiful additional evidence, postdating Mani and Johnson-Laird (1982) and Johnson-Laird (1983), that supports the notion of analog mental models as a working-memory representation. Particularly germane in the present context are the results of Byrne and Johnson-Laird (1989), who showed that deductions about spatial layouts, such as those used in the above experiments, are indeed affected by the number of possible models supported by the descriptions (i.e., the determinacy of the descriptions). This finding seems hard to explain without recourse to spatial mental models, but the reasoning tasks from which the finding derives do not implicate episodic memory.

The episodic-construction-trace hypothesis about the relationship between mental models and episodic memory accommodates all of these findings, as well as the new ones that are reported in this article. Mental models are constructed in working memory to allow inferences to be made, but the models themselves are not remembered in the longer term. What is remembered is an episodic trace of the mental operations by which mental models are constructed. Empirical evidence for this hypothesis has been

derived from two experimental manipulations. First, recognition memory for spatial descriptions is affected by the degree of trace overlap between previously seen and new descriptions (Experiments 1 and 2). Second, recognition memory is ruined when the sentences that constitute original descriptions are reordered, even when subjects are instructed to ignore sentence order (Experiment 3).

The episodic-construction-trace hypothesis is broadly related to mechanisms of human memory that have been proposed in different contexts. In proposing a special representational form for spatial mental models in working memory, the hypothesis reflects findings in the literature on working memory, in which researchers have found it necessary to distinguish a verbal, or phonological, store from a visuospatial store (see, e.g., Baddeley, 1986). In proposing that, despite the representational distinctions in working memory, long-term memory can rely on a propositional form of representation, the construction-trace hypothesis agrees with the conjecture of Glenberg and Langston (1992).

In suggesting that subjects remember aspects of the procedure by which a model is formed, rather than the model itself, the construction-trace hypothesis bears some, perhaps superficial, similarity with procedural conceptualizations of memory that derive from the view of transfer-appropriate processing (Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977). However, in the model presented here, it is a declarative trace of an encoding procedure that supports recognition, rather than reinstantiation of a procedure itself.

## REFERENCES

- ANDERSON, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, **85**, 249-277.
- ANDERSON, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- BADDELEY, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, **3**, 193-209.
- BRANSFORD, J. D., & FRANKS, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, **2**, 331-350.
- BYRNE, R. M. J., & JOHNSON-LAIRD, P. N. (1989). Spatial reasoning. *Journal of Memory & Language*, **28**, 564-575.
- DENIS, M., & CUCODE, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology*, **1**, 293-307.
- EHRlich, K., & JOHNSON-LAIRD, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning & Verbal Behavior*, **21**, 296-306.
- FRANKLIN, N., & TVERSKY, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General*, **119**, 63-76.
- FRANKLIN, N., TVERSKY, B., & COON, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, **20**, 507-518.
- GLENBERG, A. M., & LANGSTON, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory & Language*, **31**, 129-151.
- GLENBERG, A. M., MEYER, M., & LINDEM, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language*, **26**, 69-83.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.
- JOHNSON-LAIRD, P. N., & BYRNE, R. M. J. (1990). *Deduction*. Hillsdale, NJ: Erlbaum.
- KOLERS, P. A., & ROEDIGER, H. L., III (1984). Procedures of mind. *Journal of Verbal Learning & Verbal Behavior*, **23**, 425-449.
- KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- MANI, K., & JOHNSON-LAIRD, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition*, **10**, 181-187.
- MCGINN, C. (1989). *Mental content*. Oxford: Blackwell.
- MORRA, S. (1989). Developmental differences in the use of Verbatim versus spatial representations in the recall of spatial descriptions: A probabilistic model and an experimental analysis. *Journal of Memory & Language*, **28**, 37-55.
- 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.
- MORROW, D. G., BOWER, G. H., & GREENSPAN, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory & Language*, **28**, 292-312.
- MORROW, D. G., GREENSPAN, S. L., & BOWER, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory & Language*, **26**, 165-187.
- NEWELL, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- PYLYSHYN, Z. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, **80**, 1-24.
- PYLYSHYN, Z. (1979). Validating computational models: A critique of Anderson's indeterminacy of representation claim. *Psychological Review*, **86**, 383-394.
- PYLYSHYN, Z. (Ed.) (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.
- VAN DIJK, T., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

## APPENDIX A

### Example Recognition-Test Items From Experiment 1

#### Determinant Original

- The bed is to the left of the chair.
- The bed is behind the table.
- The table is to the left of the desk.
- The cupboard is in front of the table.

#### Inferrable

- The bed is to the left of the chair.
- The table is in front of the bed.
- The desk is in front of the chair.
- The table is behind the cupboard.

#### Confusion 1

- The bed is to the left of the chair.
- The bed is behind the cupboard.
- The cupboard is to the left of the table.
- The desk is in front of the cupboard.

#### Confusion 2

- The bed is to the left of the chair.
- The cupboard is in front of the bed.
- The table is in front of the chair.
- The cupboard is behind the desk.

#### Indeterminate original

- The horse is in front of the sheep.
- The cow is behind the horse.
- The sheep is to the right of the pig.
- The dog is to the left of the horse.

## Inferrable

The horse is to the right of the dog.  
 The pig is behind the dog.  
 The sheep is to the right of the pig.  
 The horse is in front of the cow.

## Confusion 1

The dog is in front of the sheep.  
 The horse is behind the dog.  
 The sheep is to the right of the pig.  
 The cow is to the left of the dog.

## Confusion 2

The dog is to the right of the cow.  
 The pig is behind the cow.  
 The sheep is to the right of the pig.  
 The dog is in front of the horse.

Note—Each item comprises an original description, an inferrable description, and two “confusion descriptions.” The four descriptions were displayed on the screen (without labels) in a random order.

**APPENDIX B****Examples of the Episodic Construction Trace for a Determinate and an Indeterminate Description**

## Determinate Original

The spoon is to the left of the knife.  
 The plate is to the right of the knife.  
 The fork is in front of the spoon.  
 The cup is in front of the knife.

## EC Trace

[start spoon knife left]  
 [plate [knife spoon] right]  
 [fork spoon front]  
 [cup knife front]

## Corresponding Inferrable Description

The spoon is to the left of the knife.  
 The plate is to the right of the knife.  
 The spoon is behind the fork.  
 The cup is to the right of the fork.

## EC Trace

[start spoon knife left]  
 [plate [knife spoon] right]  
 [fork spoon front]  
 [cup fork right]

Trace Overlap = 3

## Indeterminate Original

The sugar is in front of the ketchup.  
 The ketchup is to the right of the salt.  
 The ketchup is to the right of the pepper.  
 The salt is behind the mustard.

## EC Trace

[start sugar ketchup front]  
 [salt ketchup left]  
 [pepper ketchup left [clash pepper salt]]  
 [mustard salt front]

## Corresponding Inferrable Description

The sugar is to the right of the mustard.  
 The salt is behind the mustard.  
 The ketchup is to the right of the salt.  
 The pepper is to the left of the ketchup.

## EC Trace

[start sugar mustard right]  
 [salt mustard behind]  
 [ketchup salt right]  
 [pepper ketchup left [clash pepper salt]]

Trace Overlap = 1

## Alternative Inferrable Description for Indeterminate Original

The ketchup is to the right of the pepper.  
 The ketchup is behind the sugar.  
 The sugar is to the right of the mustard.  
 The salt is behind the mustard.

## EC Trace

[start ketchup pepper right]  
 [sugar ketchup front]  
 [mustard sugar left]  
 [salt mustard behind [clash [salt pepper]]]

Trace Overlap = 0

Note—Within each list (except the first) of each trace, the first object is the “new” object in the model. The trace overlap can vary, depending on the exact way the inferrable description is constructed. For indeterminate descriptions only, a trace overlap of 0 between original and inferrable descriptions is possible.

**APPENDIX C****Episodic Construction Traces for Sentences Used by Bransford, Barclay, and Franks (1972)**

## Old sentence 1

Three turtles rested on a floating log, and a fish swam beneath them.

## EC Trace

[start turtles log on]  
 [fish [log turtles] beneath]

## Recognition Item 1'

Three turtles rested on a floating log, and a fish swam beneath it.

## EC Trace

[start turtles log on]  
 [fish [log turtles] beneath]

Trace Overlap = 2

## Old sentence 2

Three turtles rested beside a floating log, and a fish swam beneath them.

## EC Trace

[start turtles log beside]  
 [fish turtles beneath]

Recognition Item 2'

Three turtles rested beside a floating log, and a fish swam beneath it.

EC Trace

[start turtles log beside]  
[fish log beneath]

Trace Overlap = 1

**APPENDIX D**

**Construction Trace for a Reordered Original**

Indeterminate Original

The sugar is in front of the ketchup.  
The ketchup is to the right of the salt.  
The ketchup is to the right of the pepper.  
The salt is behind the mustard.

EC Trace

[start sugar ketchup front]  
[salt ketchup left]

[pepper ketchup left [clash pepper salt]]  
[mustard salt front]

Reordered Original

The salt is behind the mustard.  
The ketchup is to the right of the salt.  
The ketchup is to the right of the pepper.  
The sugar is in front of the ketchup.

EC Trace

[start salt mustard behind]  
[ketchup salt right]  
[pepper ketchup left [clash pepper salt]]  
[sugar ketchup front]

Trace Overlap = 1

(Manuscript received August 21, 1992;  
revision accepted for publication April 5, 1993.)