

Preferred mental models in reasoning about spatial relations

GEORG JAHN

Chemnitz University of Technology, Chemnitz, Germany

MARKUS KNAUFF

Giessen University, Giessen, Germany

AND

P. N. JOHNSON-LAIRD

Princeton University, Princeton, New Jersey

The theory of mental models postulates that individuals infer that a spatial description is consistent only if they can construct a model in which all the assertions in the description are true. Individuals prefer a parsimonious representation, and so, when a description is consistent with more than one possible layout of entities on the left–right dimension, individuals in our culture prefer to construct models working from left to right. They also prefer to locate entities referred to in the same assertion as adjacent to one another in a model. And, if possible, they tend to chunk entities into a single unit in order to capture several possibilities in a single model. We report four experiments corroborating these predictions. The results shed light on the integration of relational assertions, and they show that participants exploit implicit constraints in building models of spatial relations.

Imagine that a speaker is describing a room to you. The speaker describes what is along one wall, as follows:

A table is between the TV and a chair.

Like most descriptions of a spatial relation, this assertion is compatible with many different arrangements of the entities to which it refers. Yet a theory that we will defend in this article postulates that you build a parsimonious mental model of a simple and typical spatial layout. In particular, you should be likely to imagine an arrangement in which the items of furniture are adjacent to one another and in which they are laid out in the following left-to-right order:

TV table chair.

This diagram denotes a *mental model* of the layout, but real mental models contain, not words in English, but representations of entities in the world. We use such diagrams to denote mental models throughout the present article, and for simplicity, we often refer to them as though they themselves were models. Our real concern, however, is mental models of spatial relations. Individuals have a natural tendency to construct them in order to make sense of spatial descriptions.

Early research on relational descriptions studied how individuals make inferences such as

Art is taller than Beth.

Cath is shorter than Beth.

Therefore, Art is taller than Cath.

One account of these so-called *three-term series problems* relied solely on the linguistic structure of the premises (Clark, 1969; see also Rips, 1994), and linguistic factors do affect the difficulty of these problems (Sternberg, 1981). But many results show that the relations are integrated into a linear spatial array, either a vertical one or a horizontal one (e.g., De Soto, London, & Handel, 1965; Huttenlocher, 1968; Potts & Scholz, 1975; for a review, see Evans, Newstead, & Byrne, 1993). An array can be used to derive novel conclusions or to check given conclusions even if the relation is temporal or abstract (Boroditsky, 2000; Goodwin & Johnson-Laird, 2005). The array is a mental model of the situation, and recent evidence for the use of models in reasoning about all sorts of relations comes from studies using secondary working memory tasks (Klauer, Stegmaier, & Meiser, 1997; Vandierendonck & De Vooght, 1997) and neuroimaging (e.g., Fangmeier, Knauff, Ruff, & Sloutsky, 2006; Knauff, Fangmeier, Ruff, & Johnson-Laird, 2003).

In the present article, we examine the theory of mental models and some of the consequences of its principle of parsimony. Three-term series problems, however, do not reveal much about the structure of models, and so, in order to investigate parsimony, we used a novel sort of relational

reasoning in which individuals do not draw a conclusion from premises but, instead, decide whether a set of spatial assertions is consistent or inconsistent. This task occurs spontaneously in the comprehension of everyday discourse, especially when individuals are obliged to follow instructions about, say, a stage setting, the assembly of a kit, or the directions to a destination. But it has also been used in the laboratory, both as an object of study in its own right (Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000) and as an indicator of whether or not individuals keep track of spatial relations (e.g., Jahn, 2004; O'Brien & Albrecht, 1992). In the present studies, the participants had to evaluate the consistency of spatial descriptions—for example, the following description based on the previous example:

A table is between the TV and a chair.

The light is on the left of the TV.

The table is next to the light.

If you construct the mental model of the first relation, which we illustrated above, you can add the information from the second assertion to your model:

light TV table chair.

But now the third assertion is false in this model, and you may suppose that the description as a whole is inconsistent. In fact, as we will show presently, the description *is* consistent, but it flouts the preferences that underlie the normal construction of mental models. The erroneous evaluation of inconsistency, if it were to occur, would accordingly show that individuals have constructed a model with the structure shown here.

In what follows, we will present the theory of mental models and its account of how reasoners carry out the task of assessing the consistency of descriptions. The theory yields three main predictions about performance. We then will describe four experiments designed to test these predictions. Finally, we will discuss the results and draw some general conclusions about the nature of mental models representing spatial relations.

Spatial Reasoning With Mental Models

The theory of mental models, or the *model theory* for short, postulates that individuals use the meaning of assertions and general knowledge to construct models of the possibilities compatible with assertions (Johnson-Laird, 1983). The theory applies to the integration of separate relations into models (Byrne & Johnson-Laird, 1989; Goodwin & Johnson-Laird, 2005), and it postulates that the difficulty of dealing with a relation depends on the number of arguments it concerns—that is, its *relational complexity* (Goodwin & Johnson-Laird, 2005; Halford, Wilson, & Phillips, 1998; Hummel & Holyoak, 2003). In contrast to theories based on formal rules of inference (e.g., Rips, 1994), the model theory postulates that the logical consequences of relations emerge from models: A conclusion is valid only if it holds in all the possible models of the premises.

Our present research concerns the evaluation of the consistency of spatial descriptions—that is, whether or not all

the assertions in a description can be true at the same time. According to the model theory, individuals carry out a test of consistency by trying to construct a model of a possibility in which all the assertions hold. If they succeed, they judge that the assertions are consistent; otherwise, they judge that they are inconsistent (Johnson-Laird et al., 2000). The model theory is based on a fundamental principle of *parsimony* according to which individuals tend to construct only a single, simple, and typical model (Goodwin & Johnson-Laird, 2005; Johnson-Laird, 2006). A crucial issue is, accordingly, which model of a set of assertions individuals are likely to construct first—that is, which is the preferred model (Knauff, Rauh, & Schlieder, 1995; Knauff, Rauh, Schlieder, & Strube, 1998; Rauh et al., 2005). Depending on this initial model, the task of evaluating consistency can be easy or hard. Because the evaluation of consistency mirrors the everyday comprehension of spatial descriptions, individuals need to take into account preferred models in order to ensure that their remarks are understood correctly. The process is usually automatic, because interlocutors share preferences (Garrod & Pickering, 2004). As in our opening example, speakers tend to describe a room by enumerating the furniture that it contains in a clockwise direction, working from left to right (Taylor & Tversky, 1992). When you listen to a description, you expect to proceed in this direction, and you will envisage the room accordingly. The model theory postulates that this preference is part of the process of comprehension and that it is a procedure for building a single model from a description that is compatible with multiple layouts.

The preference for working from left to right was reported informally in an early study (De Soto et al., 1965), and it reflects a cultural bias to scan in the same direction as reading and writing (Chan & Bergen, 2005; Spalek & Hammad, 2005). As another example, consider the assertion, *The lamp is next to the magazine*. Granted that the objects are laid out on the left–right dimension, there are two layouts consistent with this assertion, but the scanning bias should lead individuals to construct this model:

lamp magazine.

The tokens for the lamp and the magazine are ordered from left to right in the order in which they are referred to in the assertion. We refer to this bias as the *left-to-right* preference (see also Hörnig, Oberauer, & Weidenfeld, 2006).

If there is a left-to-right preference, the evaluation of consistency should be relatively easy when the remaining assertions can be added immediately to an existing model of the preferred sort. Some assertions, as our opening example illustrated, cannot be smoothly added to such a model. Consider, again, the description

A table is between the TV and a chair.

The light is on the left of the TV.

The table is next to the light.

The initial pair of assertions yield, as we have seen, the following model:

light TV table chair.

But, the third assertion is false in this model, and so some individuals may be tempted to respond that the three assertions are inconsistent. The response is wrong, but it is necessary to start over in order to reach the correct response. The first assertion is consistent with an alternative possibility that violates the left-to-right preference:

chair table TV.

The second assertion updates the model:

chair table light TV.

The third assertion holds in this model, and so the correct response is, *yes, the three assertions are consistent*, because they all hold in this layout. The theory accordingly predicts that problems in which a correct response depends on reordering the entities in a model should be harder than problems in which no such reordering is necessary. That is, *reordering* problems should tend to yield more errors and to take longer than control problems.

In the preceding model, the interpretation of the second assertion located the two entities to which it referred as adjacent to one another. And in general, individuals should prefer models in which entities in the same assertion are next to one another. It is less work to envisage such a model than one in which the entities are located at a distance from one another with other entities between them. Given an assertion, such as *the grape is to the right of the orange*, individuals should accordingly construct the following model:

orange grape.

A second assertion, such as *the apple is to the left of the grape*, creates a dilemma about adjacency. In one case, the adjacency of the entities in the existing model can be maintained, but then the two entities in the second assertion cannot be adjacent to one another:

apple orange grape.

In the other case, the entities in the existing model are split apart, so that they are no longer adjacent to one another, but the apple can be inserted between them, and so the entities in the second assertion are adjacent to each other:

orange apple grape.

Everyday discourse is likely to favor this second model, because the first model could be described unambiguously with a second assertion, such as *the apple is to the left of the orange*. However, when a dilemma about adjacency occurs, individuals should take longer to interpret the assertion creating it. Previous studies have used dilemmas that arise in the second assertion of a set and have shown that participants take longer to read these assertions than to read those that do not introduce a dilemma (e.g., Oberauer, Weidenfeld, & Hönig, 2006; Roberts, 2000; van der Henst & Schaeken, 2005). An open question, however, concerns the effect of the number of entities in an existing model when a dilemma occurs. As the number of such entities increases, individuals should be less likely to insert a new entity among them, because such an insertion

calls for a greater rearrangement of the model. But when the new entity is, instead, added to one end of the array, it should be easier—given the left-to-right preference—to add it to the right-hand end of the array than to the left-hand end of the array. The latter may also call for a shift of all the existing entities in the model in order to make room for one at the left-hand end.

The greater the number of entities in a model, the harder the model should be to use, but sometimes the effective number of entities can be reduced by using a *chunking* strategy (Halford et al., 1998). Consider, for example, the following description of the furniture in a room:

The TV is next to the table.

The table is to the left of the light.

The light is next to the chair.

You could try to represent all four possible orders of the items. A more parsimonious representation, however, is merely to represent two chunks, each containing a pair of items, as shown by the parentheses here:

(TV table)(light chair).

The basic principle of chunking is that items can be chunked together on a single dimension, provided that they meet three constraints: All the possible orders of the items within each chunk are consistent with the description of the layout; the order of the items within a chunk can vary independently of items outside the chunk; and the position of a chunk in the whole layout does not vary (cf. Schaeken, van der Henst, & Schroyens, 2007; Vandierendonck, Dierckx, & De Vooght, 2004). The result of chunking is to reduce the number of possibilities that need to be kept in mind. Often, in daily life, there is no need to chunk items because nothing hinges on a representation of their possible orders. The aim of chunking is to reduce the load on working memory, and so another condition under which no chunking is to be expected is a condition that reduces the need for memory—for example, by a continued visual presentation of the description. But if it is important to bear in mind these orders, chunking is one way in which to keep track of them. For example, in reasoning about tactical moves in a team sport such as soccer, the groups of backs, midfielders, and forwards can serve as chunks in ordering the players on the front-back dimension. Individuals are likely to discover the chunking strategy for themselves, especially if reasoning places a load on the processing capacity of working memory, and those problems for which the strategy is possible should be easier than those for which it is impossible.

The experiments that we report tested three predictions that we have now motivated. First, participants should have difficulty in evaluating descriptions as consistent if they call for reordering an arrangement that results from the left-to-right preference. Second, they should have difficulty in resolving dilemmas about adjacency, in which entities in one assertion or another, but not both, can be represented as adjacent to one another. Third, they should have difficulty in coping with multiple possibilities unless

they can spontaneously develop a chunking strategy and can apply it to those problems for which it is appropriate. The fundamental principle underlying all these predictions is parsimony in constructing models: Individuals aim to construct just a single, simple, and typical mental model. This single model should reflect preferences for a left-to-right ordering, adjacency, and chunking.

EXPERIMENT 1

Experiment 1 compared problems that could be solved in a smooth extension of a preferred model with problems in which the preferred model ran into an inconsistent assertion. The participants had to write down a possible horizontal arrangement consistent with three assertions or to declare that the task was impossible. Table 1 presents examples of the different sorts of problem. In all the problems, the first assertion used a *between* relation among three entities. Hence, the left-to-right preference should bias participants toward a particular model of the first assertion. In the *neutral* problems, the second and the third assertions could be integrated with the preferred model of the first assertion. In the *reordering* problems, the third assertion was inconsistent with the preferred model, and so the correct response called for the participants to construct an alternative model of the first assertion in which the entities had to be placed in a new order. If participants start with a preferred model, the third assertion in reordering problems would force them to reorder the entities. The model theory therefore predicts that the reordering problems should take longer and yield more errors than the neutral problems.

Method

Participants. Twenty-four students at the University of Tübingen served as paid participants. The mean age of the 19 women and 5 men was 23.1 years ($SD = 4.1$).

Materials. We constructed a total of 48 reasoning problems, all of which consisted of three assertions that referred to one-dimensional layouts of four objects on the left–right dimension. Table 1 presents examples of the experimental problems, and the Appendix gives a complete list of them. The terms A, B, C, and D in the table were replaced by the names of common objects—for example, *hammer*, *saw*, and *drill*. The first two assertions in the problems had four possible layouts (see Table 1). In neutral problems, the third assertion was true in the preferred model of the previous assertions—a model constructed according to the left-to-right preference. But in reordering problems, the third assertion was false in the preferred model. The experiment also contained problems that were inconsistent; that is, the three premises could not all be true, and so no possible model could satisfy them.

We constructed the problems starting with 12 initial pairs of assertions. In all pairs, one assertion referred to a *between* relation among three entities: *C between B and D*. The other assertion in the initial pair was based on *left of*, *right of*, or *next to*: *A right of C*, *C right of A*, *A left of C*, *C left of A*, *A right of D*, *D right of A*, *A left of B*, *B left of A*, *A next to B*, *B next to A*, *A next to D*, and *D next to A*. Thus, we counterbalanced the occurrence of these relations and whether the fourth object (e.g., *hammer*) was mentioned as the subject of the relation (e.g., *The hammer is to the left of the saw*) or as its object (e.g., *The saw is to the left of the hammer*). We combined each of the 12 initial pairs of assertions with four different third assertions, which produced the 48 problems that were presented to each participant. The third assertion was based on either *next to* or *left of*, and it determined whether the result was 1 of the 14 neutral problems (all

based on *next to*), 1 of the 14 reordering problems (6 *next to*, 8 *left of* for those with *next to* in the initial pair), 1 of the 12 inconsistent problems (all *next to*), or 1 of the 8 filler problems (all *next to*). Neutral problems and inconsistent problems were constructed from each initial pair; reordering problems could be constructed from only 10 of the 12 initial pairs. In neutral problems, the third assertion was always a *next to* assertion, for which the adjacent entities occurred in the preferred model in the order in which they were mentioned—for example, *C is next to D* was true in *ABCD* (in the first example in Table 1). In inconsistent problems, the third assertion either clashed directly with the *between* assertion (*B next to D*), or clashed indirectly (e.g., *B next to A*, as in the last example in Table 1). In reordering problems, the third assertion was false in the preferred models. Finally, in the filler problems, which were used to increase variety, the third assertion used *next to* and was consistent with two or three of the existing models.

The problems and instructions were in German. We constructed 24 different sets of four objects each. The four objects in a set were from a single domain and of comparable size—for example, *hammer*, *saw*, *drill*, and *pliers*. We assigned these contents four times at random to the problems and tested equal numbers of participants with each assignment.

Procedure. The participants were tested individually and carried out the experiment on a laptop computer. They were told that their task was to decide, for each of a series of descriptions, whether or not all the assertions in the description could be true of the same horizontal layout. The problems were displayed in black on white on the laptop's screen. The presentation was self-paced. Each trial began with a simultaneous presentation of the initial two assertions. When a participant pressed the space key, the third assertion replaced them. It was presented with the prompt, *Is there any layout for which all the assertions are true?* The participants responded *yes* or *no* by pressing one of the labeled response keys on the keyboard. If they responded *yes*, they wrote down the layout that they had in mind on the answer sheet, using only the initial letters of the four objects. If they responded *no*, they wrote down the word *none* on the answer sheet for layouts. The next trial started as soon as they hit the space key. Every participant evaluated each of the 48 problems, which occurred in a pseudorandom order so that there was an even distribution of inconsistent problems. Each session started with two practice trials.

Results and Discussion

In this and the subsequent experiments, a putative solution to a consistent problem was scored as correct only if the participant responded *yes* and drew a layout consistent with

Table 1
Examples of the Reasoning Problems in Experiment 1

Assertions	Possible Layouts
Neutral	
C between B and D	
A left of B	ABCD DCAB DACB ADCB
C next to D	ABCD DCAB ADCB
Reordering	
C between B and D	
A left of B	ABCD DCAB DACB ADCB
C next to A	DCAB DACB
Neutral	
C between B and D	
D next to A	BCDA BCAD DACB ADCB
B next to C	BCDA BCAD DACB ADCB
Reordering	
C between B and D	
D next to A	BCDA BCAD DACB ADCB
A left of B	DACB ADCB

Note—Layouts printed in bold are in accordance with the left-to-right preference for the interpretation of the first premise.

the premises. A putative solution to an inconsistent problem was scored as correct only if the participant responded *no* and wrote down *none* for the layout. Mean response times and mean reading times for premises were computed for correctly solved problems after the elimination of outliers with response times more than three *SDs* above the mean in each problem category (3% of the responses).

Figure 1 shows the percentages of correct responses to consistent problems and their mean latencies. As the model theory predicts, the neutral problems were solved more often than the reordering problems: The difference occurred for 19 participants, was contravened by 2 participants, and yielded three ties (binomial test, $p < .0001$). The mean reading time for the initial pair of assertions was 32.3 sec ($SE = 2.4$) and, as was expected, did not differ significantly between neutral and reordering problems (related t test, $t(21) = 6.19, p < .001$, with 21 degrees of freedom because 2 participants did not solve any reordering problems). The participants were correct for 86% of the inconsistent problems, which were easy because they involved obvious clashes of the third assertion with the *between* assertion (mean response time, 12.9 sec). The participants drew 85% correct diagrams for neutral problems (see Figure 1), and, among these diagrams, 94% of them were in accordance with the left-to-right preference. All 24 participants showed this preference ($p = .524$, binomial test). Erroneous responses to consistent problems could be either to answer that no ordering was possible (*no*) or to draw a wrong diagram. As was predicted, erroneous *no* responses were more frequent with reordering problems: 64% of the errors made with reordering problems were *no* responses, as compared with only 27% of the errors made with neutral problems.

The results corroborated the model theory's principle of parsimony—that is, individuals tend to construct a single and typical model (Goodwin & Johnson-Laird, 2005)—and the predicted consequence of a left-to-right preference. This preference made neutral problems, which were constructed in accordance with it, much easier than

problems that violated the preference and, therefore, called for a new model to be constructed in order to accommodate the third assertion in the description. The participants were accordingly both faster and more accurate for the neutral problems than for the reordering problems. In order to construct an alternative model, the participants needed access to an independent representation of the *meaning* of the initial assertions. The long reading time for these assertions may reflect the participants' need to commit the meanings of these assertions to memory. They were not always successful, however. And the construction of an alternative model took two to three times longer than coping with a preferred model in the neutral problems. As Table 1 shows, after the third assertion, the reordering problems were consistent with two models, whereas the neutral problems were consistent with three or four models. Could this factor account for the difference in difficulty between them? It is most unlikely to do so, because the participants hardly ever constructed models that did not conform to the left-to-right preference.

EXPERIMENT 2

Experiment 2 was designed to test the predictions about adjacency and chunking. Participants should tend to represent entities in the same premise as adjacent to one another, and they should tend to avoid rearranging an existing model. They should also try to treat pairs of entities as a single chunk if this strategy captures all the possibilities compatible with the description. The first prediction depended on testing problems in which dilemmas in adjacency arose. In these problems, individuals could represent entities as adjacent to one another either for one assertion or else for another assertion, but not for both assertions. Table 2 presents the problems designed to examine the preferred resolutions of dilemmas in adjacency. As the table shows, the dilemmas occurred either in the first two assertions in a problem or in the last two assertions in a problem. It is less work to rearrange a model containing two entities than one containing three entities, and so individuals should be more likely to represent entities in an assertion as adjacent and to rearrange their existing models when dilemmas arise in the first two assertions in a problem than when they arise in the last two assertions.

The model theory predicts that the problems in which chunking is possible for the entities in the initial pair of assertions should be easier than those for which it is impossible. Table 3 presents the problems used in the experiment to test this prediction. In all the problems, the participants had to keep track of the possible orders of the four entities referred to in the initial pair of assertions, because the third assertion was inconsistent with a model based on the left-to-right preference. For the *chunking* problems, the participants could form two chunks in order to capture all four possible orders of the inner pair (A and C in the examples in Table 3) and the outer pair (B and D). As the constraints on chunking require, the order of items within each chunk can vary independently from the other chunk, and the positions of the chunks in the whole layout do not vary. These problems should make reordering easy, in compari-

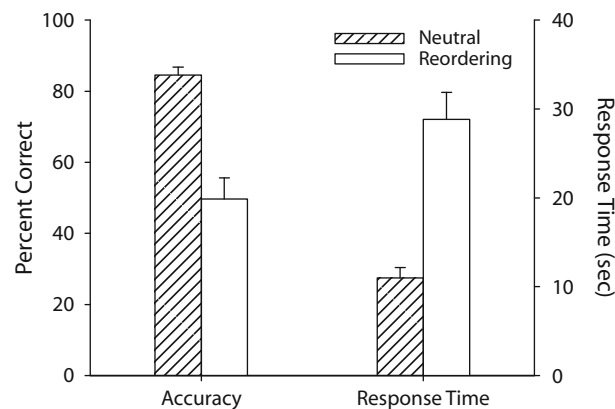


Figure 1. The percentages of correct responses to consistent problems and their mean latencies in Experiment 1 (bars show the standard errors).

Table 2
Examples of the Problems in Experiment 2 That Yielded Dilemmas About Which Two Entities Should Be Adjacent

Assertions	Possible Layouts
The dilemma about adjacency occurs in the first two assertions: B is either inserted between A and C, or else located at the left end of the array.	
C right of A B left of C D right of C	ABC BAC ABCD BACD
The dilemma about adjacency occurs in the first two assertions: B is either inserted between A and C, or else located at the right end of the array.	
A left of C B right of A D left of A	ABC ACB DABC DACB
The dilemma about adjacency occurs in the last two assertions: D is inserted between B and C, between A and B, or else located at the right end of the array.	
C right of B B right of A D left of C	ABC ABDC AD BC DABC
The dilemma about adjacency occurs in the last two assertions: D is inserted between A and B, between B and C, or else located at the right end of the array.	
B left of C A left of B D right of A	ABC AD BC ABDC ABCD

son with those problems for which chunking is impossible. In the problems for which chunking is impossible, pairs of entities remain adjacent in all possible orderings (e.g., B and C, and A and D in the second example in Table 3), but the position of such pairs in the whole layout can vary. In addition, the order within a pair does depend on the pair's position in the whole layout (B and C in the preceding example). When chunking is impossible, individuals can try to keep track of all the possible orders or they can try to retain an accurate memory for the meanings of the initial pair of assertions. Because chunking eliminates the need to lay down such memories, the reading times for the initial pair of assertions should be faster for chunking problems than for those in which chunking is impossible.

Method

Participants. Twenty-three students at the University of Tübingen who had not taken part in the first experiment served as paid participants. The mean age of the 20 women and 3 men was 24.3 years. One additional participant, who showed exceptionally poor performance (only 6% correct responses), was excluded from the analyses.

Materials. As in Experiment 1, the problems consisted of three assertions that described the left–right layouts of four objects. There were eight problems giving rise to dilemmas in adjacency. Table 2 presents examples of these problems, which gave rise to a dilemma in either the second assertion or the third assertion. In half of these problems, the critical entity to be added to the model after the dilemma arose could be inserted in the leftmost position; in half of the problems, it could be inserted in the rightmost position. There were also five reordering problems, and Table 3 presents them. Two of these problems allowed the participants to use the chunking strategy, whereas three of them did not. In all of them, the layouts that fulfilled the left-to-right preference were ruled out by the third premise. In addition, there were five inconsistent problems constructed by re-

placing the third assertion in the reordering problems with assertions that ruled out all possible layouts (*next to* assertions that directly clashed with *between* assertions).

Each of the 8 problems creating a dilemma in adjacency was presented twice, each of the 5 reordering problems was presented three times, and each of the 5 inconsistent problems was presented once. These repetitions used different contents. Hence, every participant carried out 36 problems. The same sets of objects as in Experiment 1 were used in place of A, B, C, and D. We used four random assignments of these sets to the problems and tested equal numbers of participants with each assignment.

Procedure. Experiment 2 used a procedure similar to that in Experiment 1. On each trial, the participants responded *yes* or *no* according to whether or not they inferred that there was a possible layout. If they responded *yes*, they had to draw a diagram of the layout on an answer sheet. We prepared six pseudorandom trial sequences and presented them to equal numbers of participants. There were two practice trials.

Results and Discussion

The two sorts of problems with dilemmas in adjacency did not differ reliably in accuracy (80% and 83% correct, respectively; both *SEs* = 4.6). However, the mean reading time for the initial pair of assertions was longer for problems in which the dilemma occurred within them than when it occurred in the last pair of assertions [35.6 vs. 23.8 sec, respectively; *SEs* = 3.1 and 2.0; related *t* test, $t(22) = 6.76, p < .001$]. In contrast, the response times from the onset of the third assertion were longer for problems in which it introduced the dilemma than for problems in which the second assertion introduced it [17.4 vs. 13.4 sec, respectively; *SEs* = 1.2 and 0.9; $t(22) = 4.27, p < .001$]. This predicted pattern is analogous to the results of previous studies of indeterminacy (Carreiras & Santamaria, 1997; Schaeken & Johnson-Laird, 2000; Schaeken,

Table 3
The Problems in Experiment 2 That Allowed Chunking of the Inner (Shown Within Parentheses) and Outer Pairs of Entities to Capture the Four Possibilities Compatible With the Initial Pair of Assertions, and Problems That Did Not Allow Chunking

Assertions	Possible Layouts
Chunking Possible for A and C and for B and D	
C between B and D A next to C A right of D	B(CA)D D(CA)B D(CA)B
Chunking Impossible	
C between B and D A next to D A left of B	BCAD BCDA ADCB DACB ADCB DACB
Chunking Possible for A and C and for B and D	
A between B and D C between B and D D left of B	B(AC)D D(AC)B D(AC)B
Chunking Impossible	
A between B and D A between B and C D left of B	BACD BADC DCAB CDAB DCAB CDAB
Chunking Impossible	
A between B and D C between A and D D left of B	BACD DCAB DCAB

Note—Layouts printed in bold are in accordance with the left-to-right preference.

Table 4
Percentages of Diagrams in Experiment 2 in Which the Participants Inserted an Entity Into an Existing Array, As Opposed to Adding It to One End of the Array, Which Are the Balances of Percentages in Each Cell

The sort of problem	Entity would have to be added to the left-hand end of the array	Entity would have to be added to the right-hand end of the array	Overall
The entity was in the second assertion	69	51	60
The entity was in the third assertion	58	35	47
Overall	63	43	

Note—The table shows the percentages depending on whether the entity occurred in the second or third assertion, and on whether it would have been added to the left-hand end or to the right-hand end of the array.

Johnson-Laird, & d'Ydewalle, 1996; Vandierendonck & De Vooght, 1996). However, the difference in response times for third assertions is of particular interest. In our experiment, any model consistent with the three assertions sufficed for a judgment of consistency, and so there was no need to keep track of an indeterminacy introduced by the third assertion. Yet when the third assertion introduced an adjacency dilemma, the participants took longer to respond. This phenomenon suggests that the possibility of an adjacent interpretation for either the second or the third assertion posed a dilemma to the participants that took them some time to resolve.

If the critical entity in an assertion giving rise to a dilemma is added to the end of an existing array, it has to be added either to the left end or to the right end of the array. This choice had no effect on accuracy or latency, but it did affect the diagrams that the participants drew. Table 4 presents the percentages of diagrams in which the entity was inserted into an existing array. There was one reliable effect: The participants tended to make these insertions more for entities that would have been added to the left end of an array than for entities that would have been added to the right end of an array (Wilcoxon test, $z = 3.05$, $p < .002$). Given that the participants constructed arrays from left to right, they evidently found it easier to add a new entity to the right-hand end of an array than to the left-hand end of an array, which would entail a mental shift of the array rightward.

Figure 2 shows the percentages of correct diagrams for the reordering problems, the mean reading times for the first two assertions, and the mean response times after the third assertion, depending on whether or not the chunking strategy was possible. As in the previous experiment, the proportion of erroneous *no* responses among the overall errors was higher with reordering problems than with the remaining consistent problems (67% vs. 26%). The possibility of chunking had no reliable effect on accuracy (Wilcoxon signed-rank test, two-tailed $p = .29$), but it did speed up the reading of the first two premises by 20 sec [related t test, $t(16) = 3.79$, $p = .002$], and it showed a tendency to elicit faster responses [related t test, $t(16) = 1.55$, $p = .12$]. The shorter reading times for the first two assertions suggest that chunking obviated the need to commit the meanings of these assertions to memory. They could be immediately represented in a single model in which the chunked items captured the four possible orders. The inconsistent

problems were easy, because they elicited direct clashes of *next to* and *between* assertions, and they were evaluated with an overall accuracy of 91%, with mean reading times and response times for correct solutions of 36.2 and 14.2 sec, respectively ($SEs = 2.6$ and 1.2).

The advantage of chunking entities together affected the time the participants took to read and to understand the initial pair of assertions. But trends in line with the advantage occurred in all measures. Hence, the chunking strategy should have more of an effect with problems based on a larger number of assertions, and Experiment 3 was designed to test this possibility.

EXPERIMENT 3

Entities in the initial two assertions in a description can be chunked into a single unit, provided that the resulting model captures all the possibilities compatible with these assertions. In this experiment, we increased the number of assertions in reordering problems from three to four in order to test whether the advantage of chunking would occur in accuracy with these more difficult problems. Table 5 presents examples of the problems that we used in the experiment. There were consistent problems for which chunking was possible and consistent problems

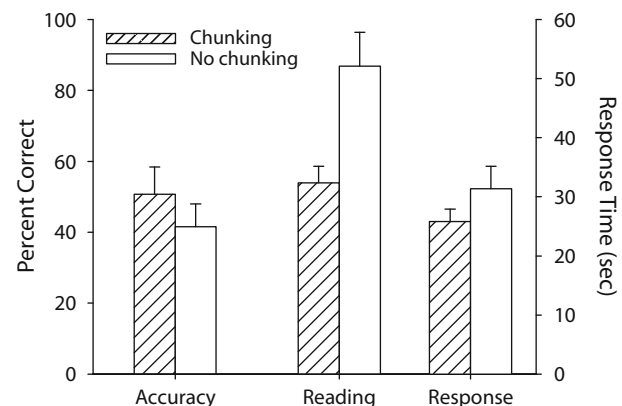


Figure 2. The percentages of correct diagrams for the reordering problems in Experiment 2, the mean reading times for the first two assertions, and the mean response times after the third assertion (bars show the standard errors), depending on whether or not chunking was possible.

for which it was impossible, and matching pairs of these problems consisted of identical assertions that differed only in the order in which the assertions occurred (see Table 5). If the participants constructed models according to the left-to-right preference, they had to revise their models after either the third or the fourth assertion.

Method

Participants. Twenty-seven Princeton undergraduates participated in the experiment. They received course credit or a small monetary compensation. The data from 3 additional participants, who disregarded the instruction not to draw a diagram before responding, were not included in the reported results.

Materials. The reasoning problems were similar to those in the previous experiments, except that each problem consisted of four assertions, instead of three, and the materials were in English. Table 5 lists examples of the problems, which were of two sorts: those in which chunking was possible and those in which it was not possible, depending on the order of the assertions. For each of the 4 problems in Table 5, a further 4 problems contained a *left of* assertion instead of the *right of* assertion. There were also 6 inconsistent problems and 4 consistent filler problems. The filler problems contained only *right of* or only *left of* assertions. In total, the 20 problems consisted of 8 reordering problems, 6 inconsistent problems, 4 filler problems, and 2 practice problems. Twenty sets of objects were assigned at random to the problems for each participant.

Procedure. The procedure was similar to that in the previous experiments. After the presentation of the two initial assertions, the third assertion was presented, and then the fourth assertion was presented with the prompt, "Is there a layout for which all [the assertions] are true?" We prepared eight different pseudorandom trial sequences and tested from 3 to 4 participants with each sequence.

Results and Discussion

The participants produced correct diagrams for 63% of the problems for which chunking was possible, but only 30% correct diagrams for the problems for which it was impossible (*SEs* = 4.6 and 5.4, respectively). Twenty-one of the participants showed this effect, 3 went against it, and there were three ties (binomial test, $p < .0001$). This effect was present for problems of the kind shown in the top half of Table 5 (54% vs. 13%, *SEs* = 7.7 and 5.0) and for problems of the kind shown in the bottom half of Table 5 (70% vs. 44%, *SEs* = 7.2 and 8.2). The accuracies for the filler problems and inconsistent problems were 72% and 88%, respectively (*SEs* = 5.6 and 2.5).

As in Experiment 2, reading the initial pair of assertions took much less time for problems in which chunking was possible (23.3 sec, *SE* = 2.0) than for problems in which it was impossible (37.4 sec, *SE* = 4.6) [related *t* test, $t(17) = 2.60, p = .02$]. Several participants solved none of the problems in which chunking was impossible, and so the degrees of freedom in the statistical tests were smaller. Response times from the onset of the fourth assertion were also marginally shorter for problems in which chunking was possible (12.9 sec, *SE* = 1.2) than for problems in which it was impossible (21.7 sec, *SE* = 3.6) [related *t* test, $t(16) = 2.04, p = .058$]. For filler problems and inconsistent problems, the mean reading times for the initial pair of assertions were 14.9 and 32.7 sec, respectively (*SEs* = 1.3 and 2.1), and mean response times were 13.2 and 12.2 sec, respectively (*SEs* = 1.1 and 1.2).

Table 5
Examples of the Problems in Experiment 3 That Allowed Chunking of the Inner (Shown in Parentheses) and Outer Pairs of Entities and the Problems That Did Not Allow Chunking

Assertions	Possible Layouts
Chunking Possible for A and C and for B and D	
C between B and D	
A next to C	B(AC)D D(AC)B
A right of D	D(AC)B
B next to C	DACB
Chunking Impossible	
C between B and D	
A right of D	BCDA DCBA DACB DCAB
B next to C	BCDA DCBA DACB
A next to C	DACB
Chunking Possible for A and C and for B and D	
A between B and D	
C between B and D	B(AC)D D(AC)B
B right of D	D(AC)B
C next to B	DACB
Chunking Impossible	
C next to B	
A between B and D	CBAD BCAD DACB DABC
B right of D	DACB DABC
C between B and D	DACB

Note—Layouts printed in bold are in accordance with the left-to-right preference.

With reordering problems consisting of four assertions, the advantage of being able to chunk entities in a model was clear in the percentages of correct diagrams, the reading times of the initial premises, and the response times. The results corroborated the reasoners' spontaneous use of the chunking strategy. It paid off, because the problems were otherwise difficult: They contained an extra premise, and so the participants had to hold in mind the information in the initial premises for a longer time.

EXPERIMENT 4

If all the assertions in a problem are presented together and the participants are allowed to view them as they sketch possible layouts, there should be no pressure to develop efficient strategies for reducing the load on working memory, and so the participants should be unlikely to develop the chunking strategy. They have no need to memorize the meanings of the initial assertions or to consider their implications ab initio, because all the assertions remain visible at the same time. We tested this prediction in Experiment 4, and it also enabled us to collect direct evidence both for the left-to-right preference and for the participants' spontaneous strategies. The same reordering problems as those in the previous experiment were evaluated, but with the assertions presented together on a single sheet of paper. They were encouraged to think aloud and to draw diagrams. We video-recorded what they had to say and what they drew. The model theory predicts that the left-to-right preference should occur in their diagrams, that the problems should be easy, and that the participants should be unlikely to develop the chunking strategy.

Method

Participants. Twenty-five Princeton undergraduates participated for course credit. None of them had taken part in the previous experiment.

Materials. The experiment used the same reordering and filler problems as those in Experiment 3 and a subset of the inconsistent problems. We prepared eight different problem sets, each consisting of eight problems, and we tested roughly equal numbers of participants with each set. The eight problems in each set were four reordering problems (with chunking possible for two of them and impossible for two of them), two filler problems, and two inconsistent problems. The selection of *right of* and *left of* variants for reordering problems was balanced across problem sets. The terms A, B, C, and D were again replaced by sets of objects. Some of the sets of objects from Experiment 3 were assigned randomly to problems in each set of problems. There were no practice problems so that we could see the very first preference in experimental problems. The problems were printed on separate sheets of paper, with the four premises toward the top of the sheet, a space for diagrams in the middle, and a line at the bottom for the participants to write their answer. A video camera was positioned above the desk and recorded the diagrams that the participants drew and their verbal comments. The purpose of this procedure was explained to the participants, who gave their consent before the start of the experiment.

Procedure. The participants were tested individually. The order of problems was pseudorandom and balanced across problem sets. The experimenter handed the problem sheets to the participants one sheet at a time. They were encouraged to draw sketches, and they were prompted to think aloud throughout each trial.

Results and Discussion

The participants performed well, as compared with the level of accuracy with reordering problems in the previous experiments, and there was no reliable difference between the problems for which chunking was possible and the problems for which it was impossible (78% correct, in each case; both $SEs = 5.8$). Similarly, there was no reliable difference in their solution times, from the presentation of the sheet of paper until the participants finished writing the solution [70.2 vs. 60.3 sec, $SEs = 6.4$ and 6.3 ; related t test, $t(22) = 1.25$, $p = .224$]. The analysis of the video recordings and the answer sheets confirmed the left-to-right preference. Initial assertions stating a *between* relation occurred in a subset of the problems, and 97 (80%) of the 121 initial diagrams for them were consistent with the left-to-right preference. The diagrams of 19 participants showed the preference, those of 5 participants did not, and there was one tie (binomial test, $p < .005$). The preference also occurred for first premises based on *next to* (71%): The diagrams of 18 participants showed the preference, those of 6 participants did not, and there was one tie (binomial test, $p < .015$).

We examined the protocols to determine the strategies that the participants had used to assess consistency. Over half the participants (14 out of 25) drew just a single layout, and 9 of them left space between the letters so that later they could insert additional entities according to the remaining assertions. The remaining participants (11) laid out all possible interpretations in diagrams for at least some of the problems. Nine of these participants also left space for filling in letters, and 3 of them also used arrows as additional annotations on their diagrams. The participants did not rely on a constant strategy; instead, they

often added corrections to a single layout or started anew. They commonly considered the assertions in the sequence in which they were written on the problem sheet, and so some instances of actual reordering could be seen when they corrected an initial diagram based on the left-to-right preference. They regularly checked their putative solution against the assertions. This checking explains the absence of any errors with inconsistent problems.

GENERAL DISCUSSION

Our experiments examined the use of mental models in the assessment of the consistency of spatial descriptions. The participants had to determine whether or not it was possible for all the assertions in a description of a spatial layout to be true (Johnson-Laird et al., 2000; Legrenzi, Girotto, & Johnson-Laird, 2003). The model theory postulates that people carry out this task by searching for a single model in which all the assertions in a description hold. If they can construct such a model, they judge that the description is consistent; otherwise, they judge that it is inconsistent. According to the theory's principle of parsimony, they tend to construct a single, simple, and typical model, even when a description is compatible with several models of alternative layouts (Goodwin & Johnson-Laird, 2005). It follows that they should focus on their preferred model and, therefore, run into difficulty if they have to switch to another possibility. The theory postulates a preference for constructing models from left to right as a result of a cultural bias in visual scanning (Chan & Bergen, 2005; Spalek & Hammad, 2005). Furthermore, participants should tend to locate entities referred to in the same assertion as adjacent to one another, and they should also avoid rearranging existing models. In general, the more entities that have to be represented in a model, the harder the task becomes. But individuals can develop strategies to reduce the load on working memory. One such strategy is to chunk together pairs of entities in a model so that they function as a single unit. This strategy allows individuals to use a single model to represent the four possible orders in which both orders of the inner and of the outer pairs are possible.

Our experiments corroborated the model theory's principle of parsimony and its predictions about model preferences. As Experiment 1 showed, the participants interpreted assertions of the form *A is between B and C* by constructing a model of B and C, working from left to right, and then inserting A between them. As a consequence, they found it easier to cope with problems that allowed them to continue with this model than with problems that forced them to construct a new model with the entities in a different spatial order. In this case, they took longer to solve the problem, and they made more errors. Often, they thought that no model was possible and thus succumbed to an illusion of inconsistency. Of course, there would have been no need to reorder entities in models if the participants had represented all the possibilities compatible with assertions or if they had had no strong left-to-right preference in interpreting assertions. Reordering, however, is impossible unless individuals

have access to a separate and independent representation of the *meaning* of assertions. The long reading times for the initial assertions in Experiment 1 may well reflect the participants' efforts to commit the meanings of these assertions to memory.

Is the left-to-right bias merely a consequence of the visual order in which terms are written in the descriptions? It does not seem so; instead, it concerns the direction in which individuals construct models. We carried out a small-scale study to test this hypothesis. The participants listened to an auditory presentation of a single spatial assertion, such as *The pen is between the candle and the stapler*. Speech has a temporal order, but not a left-to-right order. The participants had merely to listen to the assertion and then to arrange the three objects in an appropriate horizontal order. Out of 34 participants, 24 corroborated the left-to-right bias (binomial test, $p < .015$). Skeptics could argue that the preference is merely a response bias, but we doubt that individuals would respond in this way if their mental models were arranged in the opposite order. This rebuttal, of course, takes it for granted that individuals do construct mental models of spatial assertions, and so we will return to this point presently.

The model theory postulates that individuals should prefer to locate those entities that are referred to in a single assertion as adjacent to one another in a model, because this arrangement is simple and typical, in that it reflects the fact that speakers tend to describe adjacent entities when they describe a spatial layout. This preference, however, can create a dilemma. With a description, such as *The grape is to the right of the orange, the apple is to the left of the grape*, it is impossible for all three items to be adjacent to one another and still preserve their correct spatial relations. The dilemma can be resolved in one of two ways. Either the new entity—the apple—can be located appropriately at the left end of the array:

apple orange grape,

or, with some reorganization of the prior model, the new entity can be inserted between the two existing entities in the model:

orange apple grape.

The more entities that are already in a model, the more work it should take to insert a new entity among them. Experiment 2 corroborated this hypothesis. Schaeken and van der Henst (2005) have also distinguished the two ways of resolving a dilemma in adjacency. They observed that individuals were slightly more likely to resolve the dilemma by adding an entity to the end of an array (see also Ragni, Fangmeier, Webber, & Knauff, 2006). Our results, however, suggest an asymmetry: Additions of this sort are more likely to the right-hand end than to the left-hand end of an array, because of the effects of the left-to-right preference. Likewise, individuals took longer to read the initial pair of assertions when they introduced the dilemma than when the third assertion introduced it. But the response times to the third and final assertion were longer when it, rather than the second assertion, introduced the dilemma.

Experiment 2 showed that those assertions for which the chunking of adjacent entities could occur were read more quickly than those assertions for which it could not occur. In Experiment 3, we increased the number of assertions to be evaluated from three to four. The result was that the problems for which chunking could occur yielded more correct responses, shorter reading times for the initial premises, and shorter response times than did the problems for which chunking could not occur. The integration of spatial relations is known to be affected by subtle linguistic principles that we tried to counterbalance in our materials (Hörnig, Oberauer, & Weidenfeld, 2005, 2006). This control may not have been perfect, but the left-to-right preference and the tendency toward chunking were so strong that these linguistic factors cannot provide an alternative explanation for our results.

According to the model theory, individuals develop different strategies of reasoning, which reflect both differences from one individual to another and the problems on which they are working (Johnson-Laird, 2006; Roberts, 2000). We observed our participants' strategies in Experiment 4, in which all the assertions in a problem were presented together and the participants were allowed to use paper and pencil. Not surprisingly, they drew diagrams to represent the relations in the assertions. There are several ways in which individuals can keep track of alternative spatial layouts. They can try to represent all the possibilities at the same time. This task is highly demanding, and our results suggest that this is not what most of our participants did. There are strategies that place less of a load on working memory and that are consistent with the model theory. One strategy is to use a single diagram and to annotate those entities whose relative positions are indeterminate (Schaeken et al., 2007; Vandierendonck et al., 2004). This idea has been implemented in a recent computational theory of spatial reasoning according to which reasoners start with their preferred mental model, annotate those entities whose relative positions are indeterminate, and then vary this model to find alternative models that are also consistent with the assertions (Ragni, Knauff, & Nebel, 2005). Still another strategy is to build a single model and to try to remember the assertions more or less verbatim if there are other possibilities. The third author has implemented this strategy for temporal reasoning (see, e.g., Schaeken et al., 1996). The program refers back to the premises when it encounters an inconsistency with its current model. Thus, the proposed strategies that are consistent with the model theory differ with regard to whether alternative possibilities are represented via annotations or via remembered assertions in addition to a single model.

The participants in our experiments varied in the strategies that they used. When there was a load on working memory, they tended to chunk adjacent entities into single units—a strategy that reduces the *relational complexity* of problems (Halford et al., 1998). It is also similar to strategies in which reasoners annotate a single model to capture alternative possibilities (Johnson-Laird, 1983; Ragni et al., 2005; Schaeken et al., 2007; Vandierendonck et al., 2004). When all the assertions are presented at one

and the same time and are visible at all times, reasoners have much less pressure to develop efficient strategies for minimizing the load on working memory. Experiment 4 used this procedure, and the video-recording of the participants' diagrams and their thoughts as they tackled the problems provided no evidence for the chunking of adjacent items. It did reveal, however, that the participants still showed the left-to-right preference in their diagrams. More than half of them tended to draw a diagram of just a single possibility, in accordance with the principle of parsimony, although the remaining participants enumerated all the possible layouts consistent with the assertions.

So far, we have presupposed that individuals evaluate consistency by using mental models, and so readers may wonder whether there might be an alternative account for our results. The main contender is a theory based on formal rules of inference (see, e.g., Rips, 1994). Several difficulties exist for such an account. The first difficulty is that formal theories have no machinery with which to explain how individuals can draw an arrangement of entities that satisfies a spatial description. Indeed, formal theories have yet to be framed to account for the task of evaluating the consistency of a set of assertions. One formal procedure would be to try to derive a contradiction from the set of assertions; another would be to negate one of the assertions in the set and then seek to prove the resulting negation from the remaining assertions. In either case, if a proof exists, the assertions are inconsistent; otherwise, they are not. Neither of these procedures squares with what the participants in Experiment 4 had to say about how they tackled the task (see also Johnson-Laird et al., 2000).

A second difficulty for formal rules is that such theories require axioms, or *meaning postulates*, to capture the logical properties of relational terms, such as *left of* and *right of* (Hagert, 1984; van der Henst, 2002). A meaning postulate for *between* in the context of a horizontal linear ordering might be the following:

For any x , y , and z , if y is between x and z , then either x is to the left of y and z is on the right of y , or z is on the left of y and x is on the right of y .

It is hard to see how such a postulate, and reasoning based on the logical form of assertions rather than mental models, could explain the greater difficulty of problems in which individuals have to reorder entities because of a conflict with a preference for left-to-right orders. No mental representation occurs in formal theories in which such a preference could occur.

A third difficulty for formal theories is that the meaning of relational terms is highly dependent on context (Carlson & Covell, 2005; Coventry & Garrod, 2004; Miller & Johnson-Laird, 1976; van der Zee & Watson, 2005). For example, whether or not a relation such as *to the right of* is transitive depends on the tense of the assertions. If they are both in the present tense, the relation is transitive:

A is to the right of B.

B is to the right of C.

Therefore, A is to the right of C.

But if one premise is in the present tense and the other premise is in the past tense, no guarantee of transitivity exists, and individuals tend not to make the inference (Goodwin & Johnson-Laird, 2007). Dependency on context in this way is a severe problem for meaning postulates. Despite these difficulties, it may be possible to devise a formal rule theory that would explain our data. Of course, the task of evaluating the consistency of spatial descriptions may encourage individuals to rely on mental models, which they would not otherwise use in relational reasoning. However, previous results strongly support the use of models in deductive inferences from relations (see, e.g., Byrne & Johnson-Laird, 1989; Goodwin & Johnson-Laird, 2005).

We conclude that the evidence for the use of mental models in reasoning about spatial relations is substantial. It is shown in the corroboration of the need to construct a new model when an assertion violates the left-to-right preference embodied in a current model, in dilemmas about adjacency, in the chunking of entities into single units, and in the diagrams that participants draw spontaneously in order to represent assertions.

AUTHOR NOTE

Experiments 1 and 2 of this article were presented at the 27th Annual Meeting of the Cognitive Science Society; Experiments 3 and 4 were presented at the Spatial Cognition Conference, 2004. This research was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) within the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition (www.sfbtr8.uni-bremen.de) to M.K. and by NSF Grant BCS-0076287 to P.N.J.-L. to study strategies in reasoning. M.K. was also supported by a Heisenberg Award from the DFG. We thank Klaus Oberauer, Geoffrey Goodwin, Uri Hasson, Louis Lee, and three anonymous reviewers for their helpful comments. Correspondence should be addressed to G. Jahn, Department of Psychology, Chemnitz University of Technology, D-09107 Chemnitz, Germany (e-mail: georg.jahn@phil.tu-chemnitz.de).

REFERENCES

- BORODITSKY, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, **75**, 1-28.
- BYRNE, R. M. J., & JOHNSON-LAIRD, P. N. (1989). Spatial reasoning. *Journal of Memory & Language*, **28**, 564-575.
- CARLSON, L. A., & COVELL, E. (2005). Defining functional features for spatial language. In L. A. Carlson & E. van der Zee (Eds.), *Functional features in language and space* (pp. 175-190). Oxford: Oxford University Press.
- CARREIRAS, M., & SANTAMARIA, C. (1997). Reasoning about relations: Spatial and nonspatial problems. *Thinking & Reasoning*, **3**, 191-208.
- CHAN, T. T., & BERGEN, B. (2005). Writing direction influences spatial cognition. In B. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 412-417). Mahwah, NJ: Erlbaum.
- CLARK, H. H. (1969). Linguistic processes in reasoning. *Psychological Review*, **76**, 387-404.
- COVENTRY, K. R., & GARROD, S. C. (2004). *Saying, seeing, and acting: The psychological semantics of spatial prepositions*. Hove, U.K.: Psychology Press.
- DE SOTO, C. B., LONDON, M., & HANDEL, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality & Social Psychology*, **2**, 513-521.
- EVANS, J. S. B. T., NEWSTEAD, S. E., & BYRNE, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, U.K.: Erlbaum.
- FANGMEIER, T., KNAUFF, M., RUFF, C. C., & SLOUTSKY, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience*, **18**, 320-334.
- GARROD, S., & PICKERING, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, **8**, 8-11.

- GOODWIN, G. P., & JOHNSON-LAIRD, P. N. (2005). Reasoning with relations. *Psychological Review*, **112**, 468-493.
- GOODWIN, G. P., & JOHNSON-LAIRD, P. N. (2007). *Transitive and pseudo-transitive inferences*. Manuscript submitted for publication.
- HAGERT, G. (1984). Modeling mental models: Experiments in cognitive modeling of spatial reasoning. In T. O'Shea (Ed.), *Advances in artificial intelligence* (pp. 389-398). Amsterdam: North-Holland.
- HALFORD, G. S., WILSON, W. H., & PHILLIPS, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral & Brain Sciences*, **21**, 803-864.
- HÖRNIG, R., OBERAUER, K., & WEIDENFELD, A. (2005). Two principles of premise integration in spatial reasoning. *Memory & Cognition*, **33**, 131-139.
- HÖRNIG, R., OBERAUER, K., & WEIDENFELD, A. (2006). Between reasoning. *Quarterly Journal of Experimental Psychology*, **59**, 1805-1825.
- HUMMEL, J. E., & HOLYOAK, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, **110**, 220-264.
- HUTTENLOCHER, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, **75**, 550-560.
- JAHN, G. (2004). Three turtles in danger: Spontaneous construction of causally relevant spatial situation models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 969-987.
- JOHNSON-LAIRD, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.
- JOHNSON-LAIRD, P. N. (2006). *How we reason*. Oxford: Oxford University Press.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., GIROTTO, V., & LEGRENZI, M. S. (2000). Illusions in reasoning about consistency. *Science*, **288**, 531-532.
- KLAUER, K. C., STEGMAIER, R., & MEISER, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking & Reasoning*, **3**, 9-47.
- KNAUFF, M., FANGMEIER, T., RUFF, C. C., & JOHNSON-LAIRD, P. N. (2003). Reasoning, models, and images: Behavioral measures and cortical activity. *Journal of Cognitive Neuroscience*, **15**, 1-15.
- KNAUFF, M., RAUH, R., & SCHLIEDER, C. (1995). Preferred mental models in qualitative spatial reasoning: A cognitive assessment of Allen's calculus. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 200-205). Mahwah, NJ: Erlbaum.
- KNAUFF, M., RAUH, R., SCHLIEDER, C., & STRUBE, G. (1998). Mental models in spatial reasoning. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge* (Lecture Notes in Computer Science, Vol. 1404, pp. 267-291). Berlin: Springer.
- LEGRENZI, P., GIROTTO, V., & JOHNSON-LAIRD, P. N. (2003). Models of consistency. *Psychological Science*, **14**, 131-137.
- MILLER, G., & JOHNSON-LAIRD, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press, Belknap Press.
- OBERAUER, K., WEIDENFELD, A., & HÖRNIG, R. (2006). Working memory capacity and the construction of spatial mental models in comprehension and deductive reasoning. *Quarterly Journal of Experimental Psychology*, **59**, 426-447.
- O'BRIEN, E. J., & ALBRECHT, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 777-784.
- POTTS, G. R., & SCHOLZ, K. W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning & Verbal Behavior*, **14**, 439-452.
- RAGNI, M., FANGMEIER, T., WEBBER, L., & KNAUFF, M. (2006). Complexity in spatial reasoning. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1986-1991). Mahwah, NJ: Erlbaum.
- RAGNI, M., KNAUFF, M., & NEBEL, B. (2005). A computational model for spatial reasoning with mental models. In B. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1064-1070). Mahwah, NJ: Erlbaum.
- RAUH, R., HAGEN, C., KNAUFF, M., KUB, T., SCHLIEDER, C., & STRUBE, G. (2005). From preferred to alternative mental models in spatial reasoning. *Spatial Cognition & Computation*, **5**, 239-269.
- RIPS, L. J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- ROBERTS, M. J. (2000). Strategies in relational inference. *Thinking & Reasoning*, **6**, 1-26.
- SCHAEKEN, W., & JOHNSON-LAIRD, P. N. (2000). Strategies in temporal reasoning. *Thinking & Reasoning*, **6**, 193-219.
- SCHAEKEN, W., JOHNSON-LAIRD, P. N., & D'YDEWALLE, G. (1996). Mental models and temporal reasoning. *Cognition*, **60**, 205-234.
- SCHAEKEN, W., & VAN DER HENST, J. B. (2005). It's good to be wrong: An analysis of mistakes in relational reasoning. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason: Essays in honour of Paolo Legrenzi* (pp. 51-68). Hove, U.K.: Psychology Press.
- SCHAEKEN, W., VAN DER HENST, J. B., & SCHROYENS, W. (2007). The mental models theory of relational reasoning: Premises' relevance, conclusions' phrasing and cognitive economy. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. d'Ydewalle (Eds.), *The mental models theory of reasoning: Extensions and refinements* (pp. 129-150). Mahwah, NJ: Erlbaum.
- SPALEK, T. M., & HAMMAD, S. (2005). The left-to-right bias in inhibition of return is due to the direction of reading. *Psychological Science*, **16**, 15-18.
- STERNBERG, R. J. (1981). Reasoning with determinate and indeterminate linear syllogisms. *British Journal of Psychology*, **72**, 407-420.
- TAYLOR, H. A., & TVERSKY, B. (1992). Descriptions and depictions of environments. *Memory & Cognition*, **20**, 483-496.
- VAN DER HENST, J. B. (2002). Mental model theory versus the inference rule approach in relational reasoning. *Thinking & Reasoning*, **8**, 193-203.
- VAN DER HENST, J. B., & SCHAEKEN, W. (2005). The wording of conclusions in relational reasoning. *Cognition*, **97**, 1-22.
- VAN DER ZEE, E., & WATSON, M. (2005). Between space and function: How spatial and functional features determine the comprehension of between. In L. A. Carlson & E. van der Zee (Eds.), *Functional features in language and space* (pp. 113-127). Oxford: Oxford University Press.
- VANDIERENDONCK, A., & DE VOOHT, G. (1996). Evidence for mental-model-based reasoning: A comparison of reasoning with time and space concepts. *Thinking & Reasoning*, **2**, 249-272.
- VANDIERENDONCK, A., & DE VOOHT, G. (1997). Working memory constraints on linear reasoning with spatial and temporal contexts. *Quarterly Journal of Experimental Psychology*, **50A**, 803-820.
- VANDIERENDONCK, A., DIERCKX, V., & DE VOOHT, G. (2004). Mental model construction in linear reasoning: Evidence for the construction of initial annotated models. *Quarterly Journal of Experimental Psychology*, **57A**, 1369-1391.

APPENDIX
Reasoning Problems Used in Experiment 1

Among the four orderings that are consistent with the initial pair, the orderings consistent with the preferred interpretation of *C* between *B* and *D* are indicated in bold face. In parentheses after third assertions, N denotes *neutral*, R *reordering*, I *inconsistent*, and F *filler*. The subsequent numbers indicate the number of orderings consistent with all three assertions in a problem—that is, the number of orderings that remain possible after the respective third assertion.

1. Initial pair: *C* between *B* and *D*, *A* right of *C*
BCAD BCDA DCAB DCBA
Third assertions: *B* next to *C* (N, 3); *A* next to *B* (R, 2); *B* next to *D* (I, 0); *A* next to *D* (F, 2)
2. Initial pair: *C* between *B* and *D*, *C* right of *A*
BACD ABCD DACB ADCB
Third assertions: *C* next to *D* (N, 3); *D* next to *A* (R, 2); *B* next to *D* (I, 0); *B* next to *A* (F, 2)
3. Initial pair: *C* between *B* and *D*, *A* left of *C*
BACD ABCD DACB ADCB
Third assertions: *C* next to *D* (N, 3); *D* next to *A* (R, 2); *B* next to *D* (I, 0); *B* next to *A* (F, 2)
4. Initial pair: *C* between *B* and *D*, *C* left of *A*
BCAD BCDA DCAB DCBA
Third assertions: *B* next to *C* (N, 3); *A* next to *B* (R, 2); *B* next to *D* (I, 0); *A* next to *D* (F, 2)
5. Initial pair: *C* between *B* and *D*, *A* right of *D*
BCDA DACB DCAB DCBA
Third assertions: *B* next to *C* (N, 3); *C* next to *D* (N, 3); *A* next to *C* (R, 2); *B* next to *D* (I, 0)
6. Initial pair: *C* between *B* and *D*, *D* right of *A*
BCAD ABCD BACD ADCB
Third assertions: *B* next to *C* (N, 3); *B* next to *D* (I, 0); *C* next to *A* (F, 2); *D* next to *C* (F, 3)
7. Initial pair: *C* between *B* and *D*, *A* left of *B*
ABCD DACB DCAB ADCB
Third assertions: *B* next to *C* (N, 3); *C* next to *D* (N, 3); *C* next to *A* (R, 2); *B* next to *D* (I, 0)
8. Initial pair: *C* between *B* and *D*, *B* left of *A*
BACD BCDA BACD DCBA
Third assertions: *C* next to *D* (N, 3); *B* next to *D* (I, 0); *A* next to *C* (F, 2); *C* next to *B* (F, 3)
9. Initial pair: *A* next to *B*, *C* between *B* and *D*
ABCD BACD DCAB DCBA
Third assertions: *C* next to *D* (N, 4); *D* left of *B* (R, 2); *D* left of *A* (R, 2); *A* next to *D* (I, 0)
10. Initial pair: *B* next to *A*, *C* between *B* and *D*
BACD ABCD DCAB DCBA
Third assertions: *C* next to *D* (N, 4); *D* left of *B* (R, 2); *D* left of *A* (R, 2); *A* next to *D* (I, 0)
11. Initial pair: *C* between *B* and *D*, *D* next to *A*
BCDA BCAD DACB ADCB
Third assertions: *B* next to *C* (N, 4); *D* left of *B* (R, 2); *A* left of *B* (R, 2); *B* next to *A* (I, 0)
12. Initial pair: *C* between *B* and *D*, *A* next to *D*
BCAD BCDA DACB ADCB
Third assertions: *B* next to *C* (N, 4); *D* left of *B* (R, 2); *A* left of *B* (R, 2); *B* next to *A* (I, 0)