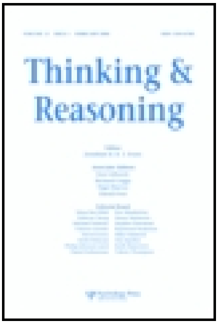
This article was downloaded by: [Princeton University]
On: 24 February 2013, At: 11:53
Publisher: Routledge
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ptar20

Strategies in temporal reasoning

Walter Schaeken [a] & Philip N. Johnson-Laird [b]

[a] University of Leuven, Belgium

[b] Princeton University, NJ, USA

Version of record first published: 24 Sep 2010.

To cite this article: Walter Schaeken & Philip N. Johnson-Laird (2000): Strategies in temporal reasoning, Thinking & Reasoning, 6:3, 193-219

To link to this article: http://dx.doi.org/10.1080/13546780050114500

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Downloaded by [Princeton University] at 11:53 24 February 2013

THINKING AND REASONING, 2000, *6* (3), 193–219

Strategies in temporal reasoning

Walter Schaeken
University of Leuven, Belgium
Philip N. Johnson-Laird
Princeton University, NJ, USA

Downloaded by [Princeton University] at 11:53 24 February 2013

This paper reports three studies of temporal reasoning. A problem of the following sort, where the letters denote common everyday events: *A happens before B. C happens before B. D happens while B. E happens while C. What is the relation between D and E?* calls for at least two alternative models to be constructed in order to give the right answer for the right reason (D happens after E). However, the first premise is irrelevant to this answer, and so if reasoners were to ignore it, then they would need to construct only one model. Experiment 1 showed that one-model problems were answered faster and more accurately than multiple-model problems. When the question preceded the premises in the statement of the multiple-model problems there was a slight tendency for the latencies of response to speed up in the predicted way. Experiment 2 modified the procedure, in part by using practice problems with many irrelevant premises, so that reasoners might grasp the advantage of ignoring them. Its results showed that when the premises preceded the question, the multiple-model problems were significantly harder than one-model problems. But when the question was presented first, the difference was significantly reduced in line with the theory's prediction. Experiment 3 used only problems with valid conclusions (i.e., one-model problems and multiple-model problems), and so the construction of multiple models was never necessary. However, there was still a significant difference between one-model problems and multiple-model problems.

How do people reason about temporal relations? For example, how do they draw the following inference:

Correspondence should be addressed to Walter Schaeken, Laboratory of Experimental Psychology, Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium, Email: Walter.Schaeken@psy.kuleuven.ac.be

Walter Schaeken was supported by the Fund for Scientific Research Flanders and IUAP/PAI P4/19. P.N. Johnson-Laird was supported in part by ARPA (CAETI) contracts N66001-94-C-6045 and N66001-95-C-8605. Part of Experiment 2 was presented to the 17th Annual Conference of the Cognitive Science Society, Pittsburgh, July 1995. We thank Maxwell Roberts and two anonymous reviewers for their helpful comments on an earlier version of the manuscript.

© 2000 Psychology Press Ltd
http://www.tandf.co.uk/journals/pp/13546783.html

Downloaded by [Princeton University] at 11:53 24 February 2013

Just after the unknown plane entered the area, there was a blip on the radar screen compatible with a surface-to-air missile.
This blip on the radar screen was just before the airliner disappeared.
Did the airliner disappear after the unknown plane entered the area?

According to the model theory, reasoners try to envisage the possible situations that would satisfy the premises, test whether the conclusion holds in these situations, and check that there is no alternative situation satisfying the premises in which the conclusion is false. If no conclusion exists to be evaluated, then reasoners can formulate one for themselves. Hence, reasoners rely on mental models, where each mental model represents what is true in a possible situation (Johnson-Laird, 1983, Johnson-Laird & Byrne, 1991; Klauer, Stegmaier, & Meiser, 1997). The premises just given elicit the following model:

unknown plane   blip   disappearance of airliner

in which "unknown plane" denotes a representation of the unknown plane entering the area, "blip" denotes a representation of the blip on the radar screen, and "disappearance of airliner" denotes a representation of the disappearance of the airliner. The model is static, and by convention time runs from left to right in these diagrams of models. The model corresponds to infinitely many different situations, e.g., it contains no explicit representation of the length of the intervals between the events. Yet all the different situations corresponding to the model have in common the truth of the two premises. The conclusion:

The airliner disappeared after the unknown plane entered the area.

is true in this model, and there is no model of the premises that refutes this conclusion. Hence, it follows validly from the premises.

We have implemented a computer program that works according to the latest version of the model theory (Schaeken, Johnson-Laird, & d'Ydewalle, 1996a). The program shows that the verbally stated theory is sufficiently coherently stated such that the claimed predictions clearly follow. We here present the computer model to elucidate the predictions of the mental model theory. Events can be described as momentary or as having durations, definite or indefinite, and as overlapping, or coinciding, in various ways (see e.g., Allen, 1983; Steedman, 1982). For simplicity, however, the program does not represent the relative durations of events: It assumes by default that they are all of roughly the same duration. It depends on three main components that respectively interpret premises, formulate conclusions, and try to refute conclusions.

The interpretation of premises depends on a "compositional semantics". Each word in the program's vocabulary, including the connectives "before", "after", and "while", has an entry that specifies its meaning, and its contribution to the truth conditions of assertions. Similarly, the program has a grammar for a small subset of English that includes such sentences as:

A happens after B

As an input sentence is parsed, the compositional semantics uses semantic rules matched to the rules in the grammar to construct a propositional representation of the sentence. This representation in effect captures the meaning (or intension) of the sentence. Thus, the propositional representation of the sentence above is:

((1 0) (A) (B))

The meaning of "after" is represented by the list (1 0), where the first parameter enables a model to be scanned in a left to right direction, that is, by incrementing this axis in unit steps of 1. Temporal models take the form of a two-dimensional array in which, by convention, time flows from left to right, and the other dimension allows for the representation of contemporaneous events. Thus, the lexical representation of "before" is (–1 0), and the representation of "while" is (0 1). The effect of these parameters is to scan the whole of the appropriate area in an array.

The propositional representation of an assertion, such as ((1 0) (A)(B)), is used to update the set of models. The program must first determine what information in the propositional representation is new, and what is old. In particular, it checks whether A or B, or both of them, is already represented in a model of the discourse. If neither of them is in an existing model, as in the case of the first sentence in a description, then the program constructs a new model:

```
b     a
```

If only one of the two referents is in an existing model, then the program adds the new event to the model according to the meaning of the temporal connective. Thus, the interpretation of the following two assertions:

A happens after B.
B happens before C.

yields models of the three alternative possibilities:

```
         (1)                 (2)                 (3)

   b      a      c      b      c      a      b      a

                                                    c
```

The third model represents A and C as happening contemporaneously.

Downloaded by [Princeton University] at 11:53 24 February 2013

The final stage of the program is to evaluate a given conclusion, or to answer a question about two events. Given a question, the program formulates a conclusion if a common relation holds between the two events over all the models of the premises, otherwise, it responds that there is no definite relation between the two events, either because different relations occur in different models or because the events do not occur together in any single model.

The main prediction of the theory is that inferences that call for only one model should be easier than those that depend on multiple models. Consider, for example the following problem:

1. a happens before b.
   b happens before c.
   d happens while b.
   e happens while c.
   What is the relation between d and e?

The program constructs a model corresponding to the following array:

a	b	c
	d	e

from which it formulates the answer that d happens before e. The premises yield only one model, and so the inference should be relatively easy. In contrast, consider the following problem, which differs from the previous one only in its second premise:

2. a happens before b.
   c happens before b.
   d happens while b.
   e happens while c.
   What is the relation between d and e?

Because of the indeterminacy created by the first two premises, the program constructs the following three models

	(1)			(2)		(3)	
a	c	b	c	a	b	a	b
	e	d	e		d	c	
						e	d

Downloaded by [Princeton University] at 11:53 24 February 2013

They each support the answer:

d happens after e

and so it is valid. In this case, however, the inference should be reliably harder because the premises yield multiple models.

Previous studies corroborated this prediction (Schaeken et al., 1996a, b; see also Carreiras & Santamaria, 1997; Schaeken, Girotto, & Johnson-Laird, 1998; Vandierendonck & De Vooght, 1996). It also confirmed that the hardest problems were those that elicited multiple models that did not support any valid answer to the question, as in the following case:

3. a happens before b.
   c happens before b.
   d happens while a.
   e happens while c,
   What is the relation between d and e?

The program constructs the following three models:

(1)			(2)			(3)	
a	c	b	c	a	b	a	b
d	e		e	d		c	
						d	
						e	

As these models show, there is no common temporal relation between d and e, and so the problem does not have a valid answer. The theory predicts that such problems should be harder than multiple-model problems with a valid answer. If reasoners construct only one of the models for a multiple-model problem, they can still formulate the correct response for a problem with a valid answer, but they will draw an erroneous conclusion for a problem with no valid answer. In other words, it is crucial to construct multiple models for the problems with no valid answers, but it is not crucial to construct them for the problems with a valid answer. Of course, when reasoners deal with the premises one by one, they cannot tell which sort of problem the premises describe until they represent the third and fourth premises, but by then they should have already constructed the multiple models required by the first two premises. The earlier studies corroborated another prediction of the model theory. As expected, reasoners took reliably longer to read the second premise of the multiple-model problems, which leads to multiple models, than the second premise of the one-model problems, which does not lead to multiple models.

Downloaded by [Princeton University] at 11:53 24 February 2013

A FUNDAMENTAL REASONING MECHANISM OR A STRATEGY?

Roberts (1993, 2000b) has argued that studies of deductive reasoning have neglected the role of strategies. Indeed, a theory of reasoning that claims to describe the processes used by all reasoners for all sorts of reasoning tasks is severely jeopardised if not all reasoners rely on the same processes. What, according to Roberts, requires much more attention is the possibility that different people use different strategies for different tasks. The role of the precise task could be substantial. This hypothesis leads to an important observation about previous studies of two-dimensional relational (spatial or temporal) reasoning.

In most previous studies (Schaeken et al., 1996b; see also Carreiras & Santamaria, 1997; Vandierendonck & De Vooght, 1996), the premises were presented one at a time followed by the question. This procedure might have influenced participants to use a "model-construction" strategy (see also Roberts, 2000b). Previous research on the three-term series partially supports this claim. Consider the following three-term series problem:

> Ann is taller than Barbara. Barbara is taller than Caroline. Who is tallest?

De Soto, London, and Handel (1965) argued that participants solved such problems using a spatial representation based on two principles. First, the principle of directional preference postulates that it is easier to build a vertical representation from top to bottom than to build one in the opposite direction. Second, the principle of end anchoring postulates that it is easier to build a representation that relates an item at one end of a spatial array to an item in the middle than to build one in the opposite direction. Huttenlocher (1968; Huttenlocher & Higgins, 1971) proposed a similar theory, but she gave an alternative explanation for the principle of end anchoring. Both these theories can be interpreted as giving support for some kind of a "model-construction" strategy. In contrast, Clark (1969, 1971) vehemently argued against these two theories, and concluded that linguistic processes (e.g., lexical marking) explained the findings. However, there was a striking difference between the experiments of the rival theorists. On the one hand, Clark (1969) used a simultaneous presentation: the premises, the question and the answer alternatives, were simultaneously visible to the participants. On the other hand, Huttenlocher (1968) used a sequential presentation in which the experimenter read aloud the premises.

Could the difference in procedure have contributed to the results? Previous research is inconclusive. Indeed, De Soto et al. (1965) argued in favour of spatial models, but used a simultaneous presentation. However, the participants had to respond within 10 seconds. Hence, it is possible that time pressure forced them to

Downloaded by [Princeton University] at 11:53 24 February 2013

Downloaded by [Princeton University] at 11:53 24 February 2013

use models. Ormrod (1979) compared different procedures. A sequential presentation yielded results more in line with the construction of models, whereas a simultaneous presentation yielded results more in line with the use of formal rules (see also Potts & Scholtz, 1975, Shaver, Pierson, & Lang, 1975; Verweij, Sijtsma, & Koops, 1999). Unfortunately, none of the results were decisive.

Most studies of two-dimensional temporal reasoning used a sequential presentation. But some experiments have used a simultaneous presentation. Their results were in accordance with the use of models, but the effects of multiple models were smaller than those that occurred with sequential presentation. In sum, experiments indicate that a natural strategy for reasoners is to construct models of the premises, but the effects of procedure—and the difference between sequential and simultaneous presentation—are not clear cut.

EXPERIMENT 1

The aim of our first experiment was to clarify the effects of sequential and simultaneous presentation. It examined the difference in difficulty between one-model and multiple-model problems, and whether it was greater with sequential presentation than with simultaneous presentation. When the premises and question are presented simultaneously, reasoners can use the question to guide their reasoning (see Wason & Johnson-Laird, 1972). This possibility has important consequences for two-dimensional temporal problems. Consider the earlier multiple-model problem:

a happens before b.
c happens before b.
d happens while b.
e happens while c.
What is the relation between d and e?

The first premise is irrelevant to the answer to the problem. Hence, if reasoners can ignore this premise, the problem no longer calls for multiple models. It depends only on one model:

```
c     b
e     d
```

It should be easier to detect that the first premise is irrelevant when the problem is presented simultaneously than when it is presented sequentially. Moreover, if the question is stated first, instead of last, the participants could adopt a strategy in which they search for a co-referential chain of premises leading from one event in the question to the other. They would again be able to ignore the irrelevant premise. Because it is this premise that calls for multiple models, they should be able to answer the question by constructing only one model. In contrast, one-model problems and problems with no valid answers do not contain an irrelevant premise, and so the mode of presentation should have no effect on performance with them.

In order to examine these questions, our first experiment compared two kinds of presentation of the problems. We decided to contrast two opposites: a condition in which a model-construction strategy was maximally advantageous and one in which it was minimally advantageous. In the sequential condition, the premises were presented one at a time at a pace controlled by the participant, and then the question was presented as the last sentence in the problem. With this procedure, it should be difficult to discover that a premise is irrelevant. Hence, the one-model problems should be easier than the multiple-model problems. In the simultaneous condition, the question was the first sentence in a simultaneous presentation with the premises below. With this procedure, it should be easy to discover that a premise is irrelevant. Reasoners should ignore it, with a consequent reduction, or disappearance, of the difference in difficulty between the one-model and multiple-model problems. However, if individuals find it impossible to ignore irrelevant information, the difference would still occur.

Downloaded by [Princeton University] at 11:53 24 February 2013

Method

Participants. A group of 32 second-year psychology students at the University of Leuven carried out the experiment as part of a course requirement.

Design. The participants acted as their own controls and carried out three sorts of problems in two different conditions. The three sorts of problem were as follows:

1. One-model problems with valid answers, which we refer to as “one-model problems”.
2. Multiple-model problems with valid answers, which we refer to as “multiple-model problems”.
3. Multiple-model problems with no valid answers, which we refer to as “problems with no valid answers”.

Table 1 presents examples of each of the three sorts of problems. In the sequential condition, the premises were presented one at a time, at a rate under the participants’ control, followed by the question about two of the events. In the simultaneous condition, the premises were presented simultaneously headed by the question. The problems in the two conditions were presented in separate blocks, one after the other. Half the participants carried out the sequential

TABLE 1
Types of problems

One-model problem	*Multiple-model problem*	*No valid answer problem*
A happens before B.	A happens before B.	A happens before B.
B happens before C.	C happens before B.	C happens before B.
D happens while A.	D happens while C.	D happens while C.
E happens while C.	E happens while B.	E happens while A.
What is the relation between between D and E?	What is the relation between D and E?	What is the relation D and E?
Mental models:		
A B C	A C B C A B A B	A C B C A B A B
D E	D E D E C	E D D E C
	D E	D

Examples of the form of the one-model, multiple-model and problems with no valid answers in Experiment 1. The table also shows the mental models for these problems.

problems and then the simultaneous problems, and half the participants carried out the two blocks in the opposite order.

The participants carried out four versions of each of the six different sorts of problems, making a total of 24 trials. The four versions were constructed in the following way: The first premise contained either the temporal relation "before" or the temporal relation "after", and the order of the two events in the question was counterbalanced so that for half of the problems the subject of the question referred to the earlier of the two events, and for half of the problems the subject of the question referred to the later of the two events. The second premise contained whichever temporal connective did not occur in the first premise. The arrangement of the events in the first pair of premises then ensured that each version expressed the same underlying meaning. All the problems had a different content, and they were presented in a different random order to each participant. The experiment measured both the accuracy and the latency of the participants' responses.

Materials. The experiment was carried out in Leuven, Belgium and the materials were in Flemish (Dutch). The content of the problems concerned everyday activities, for example

John took a shower before he washed the dishes.

Each problem was based on a separate set of activities selected at random from a pool of such events, and the names were selected at random from a pool of 24 female and 24 male first names.

Downloaded by [Princeton University] at 11:53 24 February 2013

Procedure. The participants were tested in small groups of at most four individuals, and the experiment was carried out using computers separated from each other in the same large room. The instructions were presented on the VDU. They explained that the task was to answer a series of questions based on the information in the assertions, and that the answers should be those that must be true given the truth of the previous assertions. If the participants thought that there was no definite answer, they had to type that as their response. The participants were told that they would be timed, but that they should concentrate on making the correct response.

At the beginning of each trial in the sequential condition, the screen signalled "press space-bar for the next problem". When a participant pressed the space-bar, the first premise appeared and stayed on the screen until the participant pressed the space-bar again. At this point, the next premise appeared, and the procedure continued until the fourth premise. When the participant next pressed the space-bar, the question appeared, and the participant typed an answer. The computer recorded five main latencies: the time taken to read each of the four premises, and the time from the presentation of the question until the participants began to type their answer. The instructions explained how to use the space-bar to present the next premise and how to respond.

At the beginning of each trial with a simultaneous presentation, the screen signalled "press if space-bar for the next problem". When a participant pressed the space-bar, the first problem appeared. The first sentence of the problem was the question with the four premises beneath it. The participant typed an answer underneath the problem.

There were two practice problems before the start of each block. They were all one-model problems that did not depend on inferring a transitive relation. None of the participants had any difficulty with any of these practice problems.

Downloaded by [Princeton University] at 11:53 24 February 2013

Results and discussion

We present the results in three main sections: accuracy of performance, overall latencies to respond, and reading times for the individual premises in the sequential problems.

Table 2 presents the percentages of correct responses to the different sorts of problems. Overall, the one-model problems (96%) were easier than the multiple-model problems (88%; Wilcoxon's $T = 13.5$, $n = 16$, $p < .005$). Furthermore, only three participants went against this trend. The one-model problems were also easier than the no valid answer problems (Wilcoxon's $T = 4$, $n = 28$, $p < .00005$). These two comparisons here and later in the paper are not orthogonal, but both of them were highly significant. There was no significant difference between the two groups of participants (77% correct for the group who carried out the simultaneous problems in the first block, and 82% for the group who carried out the sequential problems in the first block). And there was no

TABLE 2
Experiment 1: Correct responses

	Simultaneously presented problems			*Sequentially presented problems*		
Block	*One-model*	*Multiple-model*	*No valid answer*	*One-model*	*Multiple-model*	*No valid answer*
Simultaneously presented problems in the first block	97 (9)	91 (18)	58 (38)	92 (15)	84 (14)	39 (38)
Sequentially presented problems in the first block	100 (0)	89 (22)	69 (38)	94 (11)	86 (15)	53 (43)

The percentages of correct responses (and standard deviations in brackets) in Experiment 1.

Downloaded by [Princeton University] at 11:53 24 February 2013

significant difference in accuracy between the first block (80% correct) and the second block (79%). However, the simultaneous problems (84% correct) were significantly easier than the sequential problems (75% correct; Wilcoxon's $T = 83.5$, $n = 27$, $p < .01$). This difference presumably reflects the fact that with the simultaneous problems the reasoners had all the premises available to them until they responded, whereas with the sequential problems they had to work through the premises one at a time. Because the problems with no valid answers differ in the nature of their correct response from the other two sorts of problem, which require a conclusion, we did not include them either in the present or the subsequent analyses of interactions. We were particularly interested in whether there was a smaller difference between the one-model and multiple-model problems in the simultaneous problems than in the sequential problems. Table 2 shows no such effect. Indeed, there were no significant pairwise interactions between the three variables (type of problem, presentation condition, or order of blocks). The three-way interaction was also not significant.

Table 3 presents the overall latencies to the different sorts of problem, i.e. the sum of the times to read the premises and question, and to respond, whether correctly or incorrectly[1]. The responses in the first block were significantly faster

[1]We always analysed the times to read the premises and question irrespective of whether the participants responded correctly or incorrectly. We did so for the following reason. Reasoners can solve the multiple-model problems correctly whether they construct only one model or two models. But if they construct two models, they are more likely to err. Hence, if we had taken into account only the reading times in the case of correct responses, we would underestimate the true reading times for the multiple-model problems. They would not provide us with a correct indication of the underlying processes.

TABLE 3
Experiment 1: Overall latencies

	Simultaneously presented problems			*Sequentially presented problems*		
Block	*One-model*	*Multiple-model*	*No valid answer*	*One-model*	*Multiple-model*	*No valid answer*
Simultaneously presented problems in the first block	31.0 (11.3)	41.0 (16.9)	46.2 (16.6)	45.1 (9.3)	50.2 (16.5)	50.1 (13.6)
Sequentially presented problems in the first block	46.1 (18.2)	47.9 (16.5)	62.5 (25.7)	45.2 (9.4)	52.5 (17.2)	46.9 (15.6)

The overall latencies (in s) (and standard deviations in brackets) to read the premises and to respond in Experiment 1.

Downloaded by [Princeton University] at 11:53 24 February 2013

(43.8 s) than the responses in the second block (50.3 s; Wilcoxon's T = 151, n = 32, p < .05). We suspect that the reason for the slower responses in the second block was that the participants had to adapt to the change in procedure.

For the simultaneous presentation, the participants responded faster to the one-model problems (38.6 s) than to either the multiple-model problems (44.5 s; Wilcoxon's T = 154, n = 32, p < .05) or the problems with no valid answers (54.4 s; Wilcoxon's T = 53, n = 32, p < .00005). They also responded faster to the problems when they were in the first block (39.4 s) than when they were in the second block (52.2 s; Mann-Whitney's U = 72, $n1$ = 16, $n2$ = 16, p < .05). The difference between the one-model and multiple-model problems was smaller in the second block than in the first block, but the interaction was not significant (Mann-Whitney's U = 90, $n1$ = 16, $n2$ = 16, p < .08). The smaller increase in latency for the multiple-model problems may indicate that with practice the participants did sometimes use the question to guide their interpretation of the premises.

For the sequential presentation, the participants again responded faster to the one-model problems (45.1 s); than to either the multiple-model problems (51.3 s; Wilcoxon's T = 162, n = 32, p < .05) or the problems with no valid answer (48.5 s; Wilcoxon's T = 150, n = 32, p < .05). There were no significant differences in latencies between the problems in the first block (48.2 s) and those in the second block (48.5 S; Mann-Whitney's U = 122, $n1$ = 16, $n2$ = 16, n.s.). Likewise, as we expected, the difference between the one-model and multiple-model problems was not in the least affected by whether the problems were in the first or second block.

Table 4 presents the mean times to read the individual premises and to respond to the question for the sequential problems. Because there was no significant

TABLE 4
Experiment 1: Means for sequential problems

	One-model problems	*Multiple-model problems*	*No valid answer problems*
Premise 1	9.5(3.6)	9.5(3.9)	10.1(3.0)
Premise 2	12.5(5.8)	14.3(6.6)	14.0(6.1)
Premise 3	9.1(5.2)	8.7(2.6)	8.8(2.7)
Premise 4	7.7(2.6)	8.8(3.1)	8.0(2.4)
Question	6.4(3.3)	10.1(7.8)	7.6(5.7)
Overall	45.1	51.4	48.5

The means (in s) of reading times (and standard deviations in brackets) for the four premises, the response times, and the overall latencies, for the sequentially presented problems in Experiment 1.

Downloaded by [Princeton University] at 11:53 24 February 2013

difference between the two blocks, we combined their data. The second premise of a one-model problem calls for the continuation of the construction of the single model. In contrast, the second premise of a multiple-model problem and a problem with no valid answer calls for the construction of alternative models. As Table 4 shows, the reading times corroborated this prediction. The participants were faster to read the second premise of the one-model problems (12.5 s) than to read the second premise of either the multiple-model problems (14.3 s; Wilcoxon's T = 368, n = 32, $p < .05$) or the problems with no valid answer (14.0 s; Wilcoxon's T = 369, n = 32, $p < .05$). For the other premises, there was only one significant difference. The participants read the fourth premise for the one-model problems faster (7.7 s) than the fourth premise of the multiple-model problems (8.8 s; Wilcoxon's T = 360, n = 32, $p < .05$). One possible explanation is that when reasoners have to add information to multiple models, it takes them longer. However, this notion predicts that there should also be such a difference for the third premise. Our data showed no such difference. Finally, the participants' responses to the question in the one-model problems (6.4 s) were faster than their responses to the multiple-model problems (10.1 s; Wilcoxon's T = 404, n = 32, $p < .005$), but not significantly faster than their responses to the problems with no valid answers (7.6 s; Wilcoxon's T = 312, n = 32, $p < .19$). There may well have been a speed–accuracy trade-off in the responses to the problems with no valid answers; the participants responded quite rapidly, but were often wrong, drawing a conclusion when none was warranted.

EXPERIMENT 2

The previous experiment corroborated the predictions of the model theory. For both conditions, one-model problems were easier than multiple-model problems. Roberts (2000a) found similar findings with simultaneously and sequentially

Downloaded by [Princeton University] at 11:53 24 February 2013

presented spatial problems. In our experiment, the significant difference between one-model and multiple-model problems was, however, tiny when compared with his experiments (see also Byrne & Johnson-Laird, 1989), while there was a massive difference between the multiple-model problems and the no valid conclusion problems in our experiment. A very likely reason for the difference between Roberts' and our data is a small difference in procedure. We tell our participants in the instructions that if they think that nothing follows from the premises, they have to give that as their response. Roberts (2000a) explicitly told them that for some problems there was not enough information to answer the question. This change in procedure is likely to lead to more "nothing follows" (hence, incorrect) answers to multiple-model problems.

The effect of presenting problems simultaneously might ultimately have yielded a significant effect. In the second block of trials, these problems yielded a smaller difference in latency between the one-model and the multiple-model problems, as one would expect if the question sometimes enabled the participants to ignore the irrelevant premise in the multiple-model problems, and thereby reduce them to one-model problems. However, the interaction was not significant. Moreover, the procedures for the two sorts of problems were different. For the sequential problems, the premises were presented one at a time, but for the simultaneous problems they were presented simultaneously.

We therefore used a different design in Experiment 2. We used only simultaneous presentations, but in two contrasting conditions in blocks of trials. In the *premises-first* condition, the premises were presented first with the question underneath them. In the *question-first* condition, the question was presented first in capital letters with the premises underneath. And, in order to emphasise the value of ignoring irrelevant premises, in this condition there were two practice problems that contained many irrelevant premises. If these procedural changes were successful, then the participants should be more likely to use the question to guide their reasoning. They should ignore the irrelevant premise in the multiple-model problems, turning them into one-model problems, and thereby reducing their difficulty in comparison with the truly one-model problems. However, if it is difficult to overcome a natural strategy of dealing with all the premises in the order in which they are presented, then there would still be a difference in difficulty between one-model and multiple-model problems.

Method

Participants. A group of 32 second-year psychology students at the University of Leuven carried out the experiment as part of a course requirement. They had not previously participated in any experiment on reasoning.

Design. The participants acted as their own controls and, as in Experiment 1, carried out one-model, multiple-model, and no valid answer problems, each based on four premises. There were two blocks of such problems, one block in which they were presented with the premises first, and one block in which they were presented with the question first. Half of the participants received the two blocks in one order, and half of the participants received them in the opposite order. There were four versions of each of the three sorts of problem in a block, which were generated in the same way as in Experiment 1. All the problems had a different content. The order of the 12 problems in each block was random for each participant. The two practice problems for the premises-first block were one-model problems with a valid answer that did not require a transitive inference. The two practice problems for the question-first block contained many irrelevant premises, for example:

> What is the relation between a and b?
> b happens after a.
> c happens before b.
> b happens after d.
> e happens while a.

where a, b, c, etc. denoted common events.

Materials and procedure. The materials and the procedure were identical to those in Experiment 1.

Results

Table 5 presents the percentages of correct responses to the different sorts of problem. Overall, the one-model problems (96%) were easier than the multiple-model problems (92%; Wilcoxon's $T = 20$, $n = 14$, $p < .05$). Furthermore, only two participants showed the reverse order (multiple-model problems being easier than one-model problems). The one-model problems were also significantly easier than the problems with no valid answers (56%; Wilcoxon's $T = 3$, $n = 29$, $p < .00005$). There was no overall difference between the two orders of presentation of the blocks (79% correct for the group who carried out the question-first problems in the first block, and 84% for the group who carried out the premises-first problems in the first block; Mann Whitney's $U = 95$, $n1 = 16$, $n2 = 16$, n.s.). There was no overall difference between performance in the first block (82% correct) and the second block (81%; Wilcoxon's $T = 147$, $n = 26$, n.s.). And there was no overall difference in the percentage of correct responses to the question-first problems (84% correct) and

Downloaded by [Princeton University] at 11:53 24 February 2013

TABLE 5
Experiment 2: Correct responses

Block	*Question-first problems*			*Premises-first problems*		
	One-model	*Multiple-model*	*No valid answer*	*One-model*	*Multiple-model*	*No valid answer*
Question-first problems in the first block	98 (6)	97 (9)	50 (34)	91 (13)	86 (16)	50 (21)
Question-first problems in the second block	97 (9)	98 (6)	61 (39)	97 (9)	86 (18)	64 (24)

The percentages of correct responses (and standard deviations in brackets) to problems in Experiment 2.

Downloaded by [Princeton University] at 11:53 24 February 2013

the premises-first problems (79% correct; Wilcoxon's $T = 130$, $n = 26$, n.s.). But, as predicted, the difference between one-model problems and multiple-model problems was larger for the premises-first problems (8%) than for the question-first problems (0%; Wilcoxon's $T = 13$, $n = 13$, $p < .05$). Because the problems with no valid answers yield a different sort of response ("no valid answer") from the other two sorts of problem, we did not include them in the analysis of the interactions either in accuracy or latency. There were no other significant interactions.

Table 6 presents the total times to read the premises and question, and to respond, correctly or incorrectly, to the different sorts of problems. The one-model problems (38.5 s) were responded to significantly faster than either the multiple-model problems (43.6 s; Wilcoxon's $T = 109$, $n = 32$, $p < .005$) or the problems with no valid answers (47.8 s; Wilcoxon's $T = 41$, $n = 32$, $p < .00005$). There was no significant difference in the latencies for the two orders of presentation (43.8 s for the group with a first block of premises-first problems, and 42.8 s for the group with a first block of question-first problems, Mann-Whitney's $U = 109$, $n1 = 16$, $n2 = 16$, n.s). However, the participants did respond faster in the second block (38.1 s) than in the first block (48.4 s; Wilcoxon's $T = 68$, $n = 32$, $p < .0005$). Unlike the previous experiment, the procedure was very similar in both blocks and presumably the participants were able to capitalise on their experience to respond faster in the second block. There was no significant difference in the latencies to the question-first problems (43.3 s) and the latencies to the premises-first problems (43.3 s; Wilcoxon's $T = 240$, $n = 32$, n.s.). But, as predicted, the difference in the latencies between one-model problems and multiple-model problems was greater for the premises-first problems (14.8 s) than for the question-first problems (5.9 s;

Downloaded by [Princeton University] at 11:53 24 February 2013

TABLE 6
Experiment 2: Total latencies

	Question-first problems			Premises-first problems		
Block	*One-model*	*Multiple-model*	*No valid answer*	*One-model*	*Multiple-model*	*No valid answer*
Question-first problems in the first block	40.6 (15.4)	45.9 (17.2)	57.2 (22.7)	35.5 (8.4)	39.0 (9.8)	38.2 (11.4)
Question-first problems in the second block	37.6 (6.5)	38.2 (7.1)	40.2 (12.2)	40.1 (9.2)	51.4 (18.2)	55.5 (19.6)

Total latencies (in s) (and standard deviations in brackets) to read the premises and to respond in Experiment 2.

Wilcoxon's $T = 166$, $n = 32$, $p < .06$). The interaction between type of problem and type of block was significant (Mann-Whitney's $U = 42$, $n1 = 16$, $n2 = 16$, $p < .001$): This interaction is a consequence of the fact that problems in the second block were solved faster (37.6 s) than the problems in the first block (44.5 s).

The three-way interaction was also significant (Mann-Whitney's $U = 78$, $n1 = 16$, $n2 = 16$, $p < .05$). The advantage of one-model latencies over multiple-model latencies reduced from the first block to the second block, but the reduction was greater for question-first problems than for premises-first problems. Indeed, the difference all but disappeared the second block of trials with the question-first problems.

Discussion

A natural inferential strategy is to work through the premises in the order in which they occur. In one condition of the present experiment, however, the question occurred first (in capital letters) followed by the premises, and the participants were able to use it to guide their search through the premises for a co-referential chain interconnecting the two events in the question. This strategy allowed them to ignore the irrelevant premise in the multiple-model problems, and thus reduce these to one-model problems. As a consequence, the difference in accuracy between one-model and multiple-model problems disappeared, whereas it still occurred when the premises were presented first. The same interaction was almost significant for the latencies of response. Thus, the participants were able to adopt a strategy in which they ignored the irrelevant premise in the multiple-model problems. However the adoption and/or the correct execution of the strategy took time and practice, because there was still a difference in the total latencies between the two sorts of problem when the question-first problems occurred in the first block of trials. The moral is that the natural strategy is to work through the premises in their order of presentation, and that it takes special practice problems, the highlighting of the question, and some experience, if reasoners are to learn to ignore the irrelevant premise. One contrast between our first two experiments was that the participants were slower to respond to the second block of problems in Experiment 1, but faster to respond to them in Experiment 2. The likely explanation is that there was change in procedure from one block to the other in Experiment 1 (from simultaneous to sequential presentation or vice versa), but not in Experiment 2.

EXPERIMENT 3

In the previous experiments, it is not strictly necessary to construct multiple models for the multiple-model problems, because the correct answer emerges from any model of the premises. But it is necessary to construct all the models for the no valid conclusion problems in order to grasp that there is no definite relation between the events in the question. The participants in the experiments may have noticed the need to construct multiple models for some of the problems, and this factor might have inhibited them from ignoring the irrelevant premise in the multiple-model problems. Accordingly in Experiment 3 we presented the participants only with one-model and multiple-model problems in order to examine whether this change in the form of the problems affected their inferential strategies. We also manipulated the procedure, contrasting in two separate groups the sequential presentation and the simultaneous presentation of the problems. In both groups, the premises preceded the question.

Method

Participants. A group of 48 first-year psychology students at the University of Leuven carried out the experiment as part of a course requirement. They had not taken part in any previous study of reasoning.

Design. Each participant carried out both one-model and multiple-model problems. The *sequential* group received the premises one at a time, at a rate under the participant's control, followed by the question about two of the events. The *simultaneous* group received the premises presented simultaneously followed by the question. The problems were presented in two blocks. In both blocks, we randomly presented six one-model problems and six multiple-model problems. We constructed four versions of the two kinds of problem, by manipulating the first two premises. In a first version, both premises contained *before* (this version was presented once), in a second version both premises

Downloaded by [Princeton University] at 11:53 24 February 2013

contained *after* (this version was presented once), in the third version the first premise contained *before* and the second *after* (this version was presented twice), and in the fourth version, the first premise contained *after* and the second *before* (this version was presented twice). The manipulation of the last two sentences and the question was the same as for Experiment 1. All the problems had a different content.

Materials and procedure. The materials and procedure were identical to those in the previous experiments. So, the participants were told that their task was to answer a series of questions based on information in a set of assertions. Their answers should be those that must be true given the truth of the previous assertions. If the participants thought that there was no definite answer, then they had to type that as their response.

Results

We analyse the results in three main sections: accuracy of performance, total latencies to respond to the simultaneously presented problems, and reading times for the individual premises of the sequentially presented problems.

Table 7 presents the percentages of correct responses to the different sorts of problems. Overall, the one-model problems (89%) were easier than the multiple-model problems (77%; Wilcoxon's T = 66.5, n = 35, p < .00005). Furthermore, only seven participants showed the reverse order (multiple-model problems being easier than one-model problems). There was no difference between the two groups: the sequential presentation (82%) was not reliably harder than the simultaneous presentation (82%). As Table 7 shows, the difference between the one-model and multiple-model problems was not reliably smaller in the simultaneous condition than in the sequential condition. Indeed, there was no significant interaction between the two variables (type of problem

TABLE 7
Experiment 3: Correct responses

	Simultaneously presented problems		*Sequentially presented problems*		*Total*	
	One-model	*Multiple-model*	*One-model*	*Multiple-model*	*One-model*	*Multiple-model*
Block 1:	90(16)	78(25)	88(18)	72(25)	89	75
Block 2:	92(9)	79(19)	88(13)	80(20)	90	80
Total:	91	79	88	76	89	77

The percentages of correct responses (and standard deviations in brackets) to problems in Experiment 3.

Downloaded by [Princeton University] at 11:53 24 February 2013

and condition). However, there was a significant improvement from the first block to the second block of trials only for the multiple-model problems in the sequential presentation (72% correct in the first block versus 80% correct in the second block; Wilcoxon's $T = 35$, $n = 16$, $p < .05$). Hence, the participants in this group may have learned that it was not necessary to build more than one model of the multiple-model problems.

Table 8 presents the total times to read the premises and question, and to respond to the two sorts of problems in the simultaneous presentation. The participants responded faster to the one-model problems (42.0 s) than to the multiple-model problems (49.0 s; Wilcoxon's $T = 18$, $n = 24$, $p < .00001$). They also responded significantly faster to the problems in the second block (42.8 s) than to the problems in the first block (48.2 s; Wilcoxon's $T = 70$, $n = 24$, $p < .05$). However, the interaction was not significant: The difference between the one-model and multiple-model problems did not vary reliably from the first to the second block.

Table 9 presents the mean times to read the individual premises and to respond to the question, together with the total latencies, for the two sorts of problems in the sequential condition. The participants again responded faster to the one-model problems (50.6 s) than to the multiple-model problems (54.8 s; Wilcoxon's $T = 18$, $n = 24$, $p < .0001$). They were also faster in the second block (50.0 s) than in the first block (55.4 s; Wilcoxon's $T = 50.5$, $n = 24$, $p < .005$.). The difference between the one-model and multiple-model problems was not reliably affected by whether the problems were in the first or second block.

As Table 9 shows, the reading times corroborated the prediction that the second premise of a one-model problem calls for the continuation of the construction of the single model, whereas the second premise of a multiple-model problem calls for the construction of alternative models. The participants were faster to read the second premise of the one-model problems (11.7 s) than to read the second premise of the multiple-model problems (14.2 s; Wilcoxon's $T = 36$, $n = 24$, $p < .001$). For the other premises and the question, there were no significant differences between the two sorts of problems. However, the reading times were influenced by the block in which the premises were presented. Indeed, for the one-model problems, participants were significantly faster to read the first three premises in the second block (11.4, 12.4, 12.1 s for the first block vs. 10.7, 11.0, 9.9 s for the second block; Wilcoxon's $T = 84$, $n = 24$, $p < .05$; Wilcoxon's $T = 61$, $n = 24$, $p < .01$; instead of Wilcoxon's $T = 38$, $n = 24$, $p < .001$). For the multiple-model problems, participants were significantly faster to read the first premise and the third and fourth premise of the second block (11.7, 12.3, 10.6 s for the first block vs. 10.6, 10.7, 9.5 s for the second block; Wilcoxon's $T = 68$, $n = 24$, $p < .05$; Wilcoxon's $T = 64$, $n = 24$, $p < .01$, Wilcoxon's $T = 86.5$, $n = 24$, $p < .05$). But for the multiple-model problems, there was no significant decrease in reading times of the second premise from the first block (14.5 s) to the second block (14.0 s).

Downloaded by [Princeton University] at 11:53 24 February 2013

TABLE 8
Experiment 3: Total latencies

One-model	*Multiple model*	*Total*	
Block 1:	44.8(16.0)	51.7(19.5)	48.3
Block 2:	39.2(11.5)	46.3(13.5)	42.8
Total:	42.0	49.0	45.5

Total latencies (in s) (and standard deviations in brackets) to read the premises and to respond to the simultaneously presented problems in Experiment 3.

Downloaded by [Princeton University] at 11:53 24 February 2013

TABLE 9
Experiment 3: Means for sequential problems

		One-model problems	*Multiple-models problems*
Premise 1	Block 1	11.4(4.6)	11.7(4.5)
	Block 2	10.7(4.5)	10.6(3.3)
	Total	11.0	11.2
Premise 2	Block 1	12.4(4.2)	14.5(4.5)
	Block 2	11.0(3.9)	14.0(3.9)
	Total	11.7	14.2
Premise 3	Block 1	12.1(3.8)	12.3(4.2)
	Block 2	9.9(2.5)	10.7(3.2))
	Total	11.0	11.5
Premise 4	Block 1	10.5(4.1)	10.6(3.9)
	Block 2	9.4(2.6)	9.5(2.7)
	Total	9.9	10.0
Question	Block 1	7.1(3.4)	8.2(3.0)
	Block 2	6.7(1.8)	7.6(2.5)
	Total	6.9	7.9
Overall	Block 1	53.4	57.3
	Block 2	47.7	52.3
	Total	50.6	54.8

The means (in s) of the reading times (and standard deviations in brackets) for each of the two blocks and in total for the four premises, the times to answer the question, and the overall times, for the sequentially presented problems in Experiment 3.

Discussion

The results confirm the main finding of the previous two experiments. Reasoners have a natural strategy of working through the premises in their order of presentation. They tend to construct multiple models of premises even when it is not necessary to do so in order to reach the correct conclusion. Hence, the multiple-model problems were more difficult and required more time than the one-model problems. This observation was made for both sequentially presented problems and simultaneously presented problems. There was no effect of presentation on the relative difficulty of one-model and multiple-model problems. However, when the multiple-model problems were presented sequentially, they did become easier in the second block. According to the model theory, the second premise is crucial for multiple-model problems. The participants read all the premises, apart from the second one, faster in the second block. A possible explanation is that after some experience, reasoners realised that they did not have to construct more than one model. However, the construction of multiple models is an almost automatic process, so it must be suppressed when the premise is encountered that makes the construction of multiple models possible. Hence, the faster reading times in the second block are counteracted for the second premise by the need to suppress the construction of more than one model. Another possibility is that there was a trade-off between speed and accuracy. Practice yielded faster reading times except for the second premise. The relatively longer reading enabled reasoners to construct the multiple models more carefully, and so they were more accurate on these problems than in the first block.

GENERAL DISCUSSION

What have we learned from our experiments? Certainly, the results show that inferences about temporal relations that call for just one model are significantly easier than those that call for multiple models. Reasoners drew a greater percentage of correct conclusions to the one-model problems, and they also responded faster to them. These results corroborate the central prediction of the model theory (see also Schaeken et al., 1996a). Moreover, when the premises were presented one at a time, reasoners spent longer in reading - and presumably in understanding—those premises that call for the construction of multiple models in comparison with the corresponding premises in one-model problems.

The effects of the different presentations were somewhat surprising. A simultaneous presentation of the premises did not automatically make it easier for reasoners to ignore the irrelevant premise. Roberts (2000a) has obtained a similar result. He even observed an increase in multiple-model constructions with a simultaneous presentation. We did not replicate that finding. One reason for this failure to replicate might be a difference between temporal and spatial problems. Another more likely reason, however, might be some small

Downloaded by [Princeton University] at 11:53 24 February 2013

differences in procedure. As we said before, Roberts (2000a) explicitly told the participants that for some problems there was not enough information to answer the question, whereas we did not give this information.

The results of our three experiments show that reasoners naturally adopt a strategy of treating the premises in their order of presentation, and so they build multiple models even on the basis of premises that are irrelevant to their inferential task. Although the simultaneous problems were as a whole easier than the sequential problems, the advantage of one-model over multiple-model problems occurred in both conditions. However, in all of our experiments there were signs that reasoners can adopt another strategy. In Experiment 1, the participants' overall latencies increased substantially except for the multiple-model problems when the simultaneous problems occurred in the second block of trials. Experiment 2 highlighted the importance of the question and the possible presence of irrelevant premises. Yet the participants were faster to respond to the one-model problems than to the multiple-model problems, and there was no difference between the question-first and the premises-first problems. However, the difference in accuracy between one-model and multiple-model problems disappeared when the question occurred first. In Experiment 3, there was no need to build multiple models, but one-model problems were easier than multiple-model problems in both presentation conditions. However, the multiple-model problems were significantly easier in the second block when they were presented sequentially. In other words, reasoners naturally construct multiple models, but they can develop a strategy in which they ignore the irrelevant premise in multiple-model problems. This strategy reduces the multiple-model problems to one-model problems, and so the difference in difficulty between the two sorts of problems reduces or disappears. Our results show that it takes time and practice to learn to use this strategy.

Evans, Handley, and Buck (1998) introduced a somewhat similar manipulation into conditional reasoning problems. In one experiment, they presented a conclusion (which had to be evaluated) either before or after the premises. They observed that the order of information had a significant effect on the mental processing of conditional inference problems. Although overall reasoners did not make more Modus Tollens and Denial of the Antecedent inferences when the conclusion was presented first, they made more conditional inferences with these problems when the major premise contained a negation in the inferential clause (i.e., the clause about which one is invited to make an inference). Interestingly, they observed that the participants reasoned more quickly with problems in the standard order (first the premises, next the conclusion) than in the reversed order (first the conclusion, next the premises). Evans et al. (1998) suggest that the preferred mode of reasoning is from premises to conclusion. We observed a similar effect. Presenting the question before the premises can make reasoning easier, but it takes time and practice before reasoners can really use the advantage.

Downloaded by [Princeton University] at 11:53 24 February 2013

Downloaded by [Princeton University] at 11:53 24 February 2013

The results of the third experiment can be interpreted in another way. Indeed, the fact that multiple-model problems were still more difficult, although the no valid conclusion problems were removed and hence the need to build multiple models, can be interpreted as support for the thesis that for simple problems building multiple models is the natural thing to do. Some recent evidence suggests that reasoners "satisfice", that is, they construct only one model of the premises in syllogistic reasoning (Evans, Handley, Harper, & Johnson-Laird, 1999). On the basis of this and other evidence, Evans (2000) argued that the construction of alternative models is a strategic choice, and the natural strategy is to consider just one state of affairs (mental model) at a time. The particular model is derived from tacit pragmatic processes. Granted that such a model seems satisfactory, reasoners based their inferences on them. To achieve deductive accuracy (what Evans refers to as "rationality$_2$"), one needs explicit strategic thinking, and a search for alternative models. The results of the present experiments contrast with this claim. Reasoners seem naturally to construct multiple models, and it was difficult for them to learn to ignore the irrelevant premises that necessitated these models. One possibility is that the difference arises because the relational reasoning in our experiments is easier than syllogistic reasoning. People may have difficulty in grasping the different possibilities implied by syllogistic premises. Yet, when reasoners, construct external models, they do construct multiple models even in syllogistic reasoning (Bucciarelli & Johnson-Laird, 1999).

Is there any alternative explanation for our results? One possibility is that the participants reasoned, not by using mental models, but by using formal rules of inference like those of a logical calculus (see e.g., Braine & O'Brien, 1991; Rips, 1994). These systems might be based on so-called "tense logic" (Prior, 1967; Rescher & Urquhart, 1971), but it is more plausible to use meaning postulates to capture the logical properties of temporal connectives, for example:

If x happens after y, and y happens after z, then x happens after z.

However, in Experiment 1, the one-model problems have a longer formal derivation that includes the derivation needed for the multiple-model problems. In particular, the one-model problem in Table 1 calls for a derivation to establish the relation between a and c, and then a further derivation that uses this relation to interrelate d and e, whereas the multiple model problem calls only for a derivation that interrelates d and e in virtue of the relation between b and c that is stated in the second premise. Proponents of rule theories can argue that the second problem is harder because, despite its shorter derivation, it has an irrelevant premise (see Rips, 1994). But, even when reasoners can ignore the irrelevant premise (in Experiment 2), the difference between the two sorts of problems does not reverse as it ought to according to the formal rule theories. Similarly, a recent study of spatial and temporal inferences has shown that even when all the problems have irrelevant premises, one-model problems remain easier than multiple-model problems (Schaeken, Girotto, & Johnson-Laird, 1998). Moreover, the greater reading times for the premises that call for the construction of multiple models, which we observed in Experiments 1 and 3, cannot be reconciled with theories that rely on formal rules, because the rules do not concern models, but only the derivation of one expression from another.

According to the mental model theory, and its computer implementation, people normally attempt to construct all possible models of the premises. However the process is, in principle, intractable, because the number of models grows exponentially as indeterminacies in the premises mount up. The strategy is feasible only if there is a small number of indeterminacies, and evidence exists that reasoners do try to keep track of alternative models, at least in those sorts of reasoning that do not depend on quantifiers, e.g., in reasoning that hinges only on simple relations (for a review, see Evans, Newstead, & Byrne, 1993). For more complex reasoning, multiple model construction might not be automatic (see, e.g., Evans, et al., 1999). Hence, a simple mechanism was implemented in the computer implementation of the model theory. When the number of models exceeds a certain constant representing the capacity of working memory, the program searches for a chain of premises interrelating the two events in the question, and constructs models only from this chain. For example, given the following premises:

a happens before b.
x happens before b.
y happens after z.
b happens before c.
What's the relation between a and c?

the program would construct seven alternative models of the premises. This number almost certainly exceeds the processing capacity of human working memory. Hence, if the program's constant is set for three models, then it adopts an alternative strategy as soon as it tries to interpret the third premise. It searches depth-first for a chain of premises interrelating the two events in the question, and constructs models only for them. With the example above, it discovers the following co-referential chain of premises:

a happens before b.
b happens before c.

This chain interrelates the two events in the question, a and c, and the program constructs the single model that these premises support. This model yields the valid answer:

a happens before c.

Downloaded by [Princeton University] at 11:53 24 February 2013

This strategy has the advantage that it drops all premises that are irrelevant to the inference. It also deals with the premises in a co-referential order in which each premise after the first refers to an event already represented in the set of models. In sum, the different strategies adopted by the participants in the present experiments are compatible with the basic framework of the model theory.

In conclusion, the evidence suggests that the majority of logically untrained individuals spontaneously reason about temporal relations by constructing mental models of the premises. As a result, if an irrelevant premise calls for the construction of multiple models, the task takes longer and is more liable to error. This natural strategy can be overcome by a practical demonstration of the difficulties caused by irrelevant premises and by stating the question at the head of the premises. In this case, the participants can ignore the irrelevant premise, the problem becomes a one-model problem, and is no longer any harder than the regular problems that call for one model.

Manuscript received 13 October 1997
Revised manuscript received 11 February 2000

Downloaded by [Princeton University] at 11:53 24 February 2013

REFERENCES

Allen, J. (1983). Maintaining knowledge about temporal intervals. *Communications of the Association for Computing Machinery*, *26*, 832–843.

Braine, M.D.S., & O'Brien, D.P. (1991). A theory of If. A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, *98*, 182–203.

Bucciarelli, M., & Johnson-Laird, P.N. (1999) Strategies in syllogistic reasoning. *Cognitive Science*, *23*, 247–303.

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

Carreiras, M., & Santamaria, C. (1997). Reasoning about relations: Spatial and nonspatial problems. *Thinking and Reasoning*, *3*, 191–208.

Clark, H.H. (1969). Linguistic processes in reasoning. *Psychological Review*, *76,* 387–404.

Clark, H.H. (1971). More about "adjectives, comparatives, and syllogisms": A reply to Huttenlocher and Higgins. *Psychological Review*, *78*, 505–514.

De Soto, C.B., London, M., & Handel, S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, *2*, 513–521.

Evans, J.St.B.T. (2000). What could and could not be a strategy in reasoning. In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 1–22). Mahwah, NJ: Erlbaum.

Evans, J.St.B.T., Handley, S.J., & Buck, E. (1998). Ordering of information in conditional reasoning. *British Journal of Psychology*, *89*, 383–403.

Evans, J.St.B.T., Handley, S.J., Harper, C.N.J., & Johnson-Laird, P.N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1495–1513.

Evans, J.St.B.T., Newstead, S.E, & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review*, 75, 550–560.

Huttenlocher, J., & Higgins, E.T. (1971). Adjectives, comparatives, and syllogisms. *Psychological Review*, *78*, 505–514.

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

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Klauer, K.C., Stegmaier, R., & Meiser, T. (1997). Working memory involvement in propositional and spatial reasoning. *Thinking and Reasoning*, *3*, 9–47.

Ormrod, J.E. (1979). Cognitive processes in the solution of three-term series problems. *American Journal of Psychology*, *92*, 235–255.

Potts, G.R., & Scholtz, K.W. (1975). The internal representation of a three-term series problem. *Journal of Verbal Learning and Verbal Behavior*, *14*, 439–452.

Prior, A.N. (1967). *Past, present, and future*. Oxford: Clarendon Press.

Rescher, N., & Urquhart, A. (1971). *Temporal logic*. New York: Springer-Verlag.

Rips, L.J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Roberts, M.J. (1993). Human reasoning: Deduction rules or mental models, or both? *Quarterly Journal of Experimental Psychology*, *46A*, 569–589.

Roberts, M.J. (2000a). Strategies in relational inference. *Thinking and Reasoning*, *6*, 1–26.

Roberts, M.J. (2000b). Individual differences in reasoning strategies: A problem to solve or an opportunity to seize? In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Deductive reasoning and strategies* (pp. 23–48). Mahwah, NJ: Erlbaum.

Schaeken, W., Girotto, V., & Johnson-Laird, P.N. (1998). The effect of an irrelevant premise on spatial and temporal reasoning. *Kognitionswisschenschaft* , *7*, 27–32.

Schaeken, W, Johnson-Laird, P.N., & d'Ydewalle, G. (1996a). Mental models and temporal reasoning. *Cognition*, *60*, 205–234.

Schaeken, W., Johnson-Laird, P.N., & d'Ydewalle, G. (1996b). Tense, aspect, and temporal reasoning. *Thinking and Reasoning*, *2*, 309–327.

Shaver, P., Pierson, L., & Lang, S. (1975). Converging evidence for the functional significance of imagery in problem solving. *Cognition*, *3*, 359–375.

Steedman, M.J. (1982). Reference to past time. In R.J. Jarvella & W. Klein (Eds.), *Speech, place, and action*. London: Wiley.

Vandierendonck, A., & De Vooght, G. (1996) Evidence for mental-model based reasoning: A comparison for reasoning with time and space concepts. *Thinking and Reasoning*, *2*, 249–272.

Verweij, A.C., Sijtsma, K., & Koops, W. (1999). An ordinal scale for transitive reasoning by means of a deductive strategy. *International Journal of Behavioral Development*, *23*, 241–264.

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content.* London: Batsford.

Downloaded by [Princeton University] at 11:53 24 February 2013