

Tense, Aspect, and Temporal Reasoning

W. Schaeken

Laboratory of Experimental Psychology, Leuven, Belgium

P.N. Johnson-Laird

Princeton University, USA

G. d'Ydewalle

Laboratory of Experimental Psychology, Leuven, Belgium

We report two experiments on temporal reasoning with problems, such as:

John has cleaned the house.

John is taking a shower.

John is going to read the paper.

Mary always does the dishes when John cleans the house.

Mary always drinks her coffee when John reads the paper.

What for Mary is the relation between doing the dishes and drinking coffee?

The experiments showed that problems such as this one, which require one mental model, elicited correct answers more often than did those requiring multiple models (e.g. with the second premise modified to “John has taken a shower”, so that the order between the events in the first two premises is not fixed). These multiple-model problems, in turn, elicited more correct answers than did multiple-model problems with no valid answers (e.g. with the second premise modified to “John has taken a shower”, and the fifth premise modified to “Mary always drinks her coffee when John takes a shower”). One-model problems were also solved more quickly than multiple-model problems, which were solved more quickly than problems with no valid answers. These results corroborated the predictions of the mental model theory of reasoning.

INTRODUCTION

Reasoning about temporal relations is an important aspect of everyday thinking. If you know the following facts:

Requests for reprints should be sent to Walter Schaeken, Laboratory of Experimental Psychology, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium.

We thank Andreas de Troy for implementing the program used to carry out the Experiment. Schaeken and d'Ydewalle are supported by the Belgian Program on Inter-University Poles of Attraction Convention Number 31, and Johnson-Laird's research is supported in part by the John S. McDonnell foundation.

The children became ill after going to the cinema.
They ate seafood before they went to the cinema.

then you can infer:

The children became ill after they ate seafood.

Although logicians have devised various “temporal logics”, and linguists have investigated many aspects of temporal language, the topic of temporal reasoning has been strangely neglected by psychologists. They have studied the perception of time (e.g. Fraisse, 1963), how children conceive of time (e.g. Piaget, 1969), but they have carried out few studies, if any, of how individuals reason on the basis of temporal relations.

Following McTaggart (1927; see Miller & Johnson-Laird, 1976, Sec.6.2), we distinguish two ways of thinking about (and describing) temporal order. First, an event can be past, present, or future (McTaggart’s A-series). Second, one event can be earlier or later than another (McTaggart’s B-series). Any relation described in terms of the A-series can be described in terms of the B-series, but the converse is not true. We have previously made an experimental investigation of how subjects reason on the basis of B-series relations, i.e. temporal connectives, such as “before”, “after”, and “while”. The experiments were carried out in Flemish, which has connectives that correspond to these English terms (see Schaeken, Johnson-Laird, & d’Ydewalle, in press). They were designed to test the predictions of the theory of reasoning based on mental models (see e.g. Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991). The subjects in the experiments responded to three sorts of problems. The first sort is illustrated by the following premises:

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

where “a”, “b”, “c”, etc. denote descriptions of common everyday events, e.g. John drinks his coffee. According to the model theory, this problem calls for just one model:

a	b	c
d		e

where the horizontal axis represents time, and the vertical axis is used to represent contemporaneous events. The model yields the answer:

d happens before e.

Such “one-model” problems should be relatively easy.

In contrast, consider an example of the second sort of problem:

a happens before c.

b happens before c.

d happens while a.

e happens while c.

What is the relation between d and e?

These problems do not establish a definite relation between events a and b, and so the premises call for the following models representing alternative possibilities:

a	b	c
d		e

and:

b	a	c
	d	e

The premises are also consistent with the following model:

a	c
b	
d	e

but, for simplicity, we will ignore these cases in which three or more events are contemporaneous (subjects tend to ignore these cases, too). This “multiple-model” problem, however, supports the same answer, which holds in all models:

d happens before e.

It should be harder to answer this problem than the one-model problem, because of the additional load on working memory. Readers should note, however, that even if subjects construct only one of the models, they should still make the correct answer, because it holds for all of them.

Finally, consider the third sort of problem:

a happens before c.

b happens before c.

d happens while a.

e happens while b.

What is the relation between d and e?

Again, the premises leave open the relation between events b and c, but in this case the indeterminacy matters. The premises support the following models:

a	b	c
d	e	

and:

b	a	c
e	d	

In this case, there is no valid answer. This “no-valid-answer” problem should be the hardest of all, because reasoners have to construct at least two models, and to take into account the contrasting relations between events e and d in them, in order to make the correct response that there is no valid answer. For simplicity, we will henceforth refer to multiple-model problems with a valid answer as “multiple-model problems” and to multiple-model problems with no valid answer as “no-valid-answer problems”; readers should bear in mind that both these sorts of problems depend on multiple models. In general, the model theory predicts that more models mean more work—it will take reasoners longer to draw a conclusion, and they will be more likely to err.

Our experiments confirmed the predictions of the model theory (see Schaeken et al., in press). Of particular interest was an experiment in which we recorded the correctness of the responses, their latencies, and the reading times of the premises. The results corroborated the model theory’s prediction about the latencies of responses: one-model problems were solved more quickly than multiple-model problems, which were solved more quickly than problems with no valid answers. The results also confirmed another of the theory’s predictions: subjects took longer to read the premise that requires the initial construction of multiple models than the corresponding premise of the one-model problems.

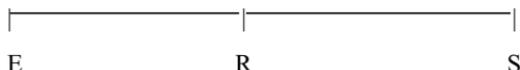
Reasoning on the basis of the B-series is akin to spatial reasoning, and the terms “before” and “after” are systematically ambiguous, referring either to spatial or to temporal relations, although the content of the clauses they interrelate usually makes clear the intended senses. Our studies of reasoning based on the B-series corroborate an earlier investigation of spatial reasoning (see Byrne & Johnson-Laird, 1989). Indeed, a friendly critic has queried whether there is anything distinctive to temporal reasoning about these studies (Martin Braine, personal communication). In contrast, the present paper reports a study of relations that are unique to temporal reasoning—those established by tense and aspect (the A-series).

Our aim is to investigate the predictions of the model theory about reasoning based on tense and aspect. English (and Flemish, the language in which the experiment was carried out) has just two tenses, the past and the present, which are indicated by a morphological change to the verbs, e.g. “loves” (present) vs. “loved” (past). Events in the future can be referred to by the present tense, e.g. “he leaves tomorrow”, or by the use of modal auxiliaries and other constructions, e.g. “he will leave tomorrow”. The description of temporal events is also handled by an aspectual system that depends on auxiliaries and other morphemes, e.g. “he is taking a shower”, which conveys an event that is ongoing in relation to the time of the speech utterance. Following Reichenbach’s (1947) seminal analysis, tense and aspect in English call for a distinction to be drawn between three sorts of points in time: the time of the speech utterance (S), the time of the event (E), and the reference time (R). These three points are distinct in the case of an utterance in the past perfect, such as:

Mary had seen the house

which can be represented by the following diagram in which the direction of time runs from left to right:

Mary sees the house



The time of the speech utterance is clearly the time at which the speaker utters the sentence. The reference time is implicit in this utterance, but it is clearly prior to the speech utterance time and after the time of the event, i.e. the time at which Mary saw the house. The reference time can be made explicit, as in the following example:

By 3 p.m. on December 4th 1901, Mary had seen the house.

The inferences that we have investigated depend on the fact that the temporal order of events can be conveyed by tense and aspect. Consider, for example, the following assertions:

John has cleaned the house.

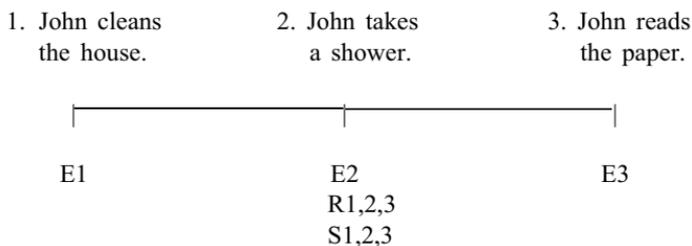
He is taking a shower.

He is going to read the paper.

Mary always does the dishes when John cleans the house.

She always drinks her coffee when he reads the paper.

The first three assertions establish the following relations (following Reichenbach's scheme):



The speech time of all three utterances is the present, and so S1, S2, and S3 coincide with E2 (the event in the present tense). Likewise, in Reichenbach's analysis, the reference time coincides with the speech time for these three simple assertions. The fourth and fifth premises add the further information about when Mary does the dishes and about when she drinks her coffee. We will represent all five premises in a diagram that corresponds to a model of their temporal relations, on the plausible assumption that the events do not significantly overlap each other in time:

cleaning dishes	shower	paper coffee
--------------------	--------	-----------------

"Cleaning" denotes John cleaning the house, "shower" denotes John taking a shower, "paper" denotes John reading the paper, "dishes" denotes Mary doing the dishes, and "coffee" denotes Mary drinking her coffee. This model supports the conclusion:

Mary does the dishes before she drinks her coffee.

This conclusion follows validly from the premises because there is no model of them in which it is false.

EXPERIMENT 1

Introduction

In our experiment, we examined three sorts of problems. The first sort of problems were one-model ones such as the example given earlier. The second sort were multiple-model problems created by using two premises that had the same tense and aspect. Consider the following example in which the first two premises are both in the present perfect:

John has cleaned the house.
John has taken a shower.

John is reading the paper.

Mary always does the dishes when John cleans the house.

Mary always drinks coffee when John reads the paper.

What for Mary is the relation between doing the dishes and drinking coffee?

These premises do not fix the temporal relation between John cleaning the house and taking a shower, and so they are consistent with multiple models:

Cleaning Dishes	Shower	Paper Coffee
--------------------	--------	-----------------

and:

Shower	Cleaning Dishes	Paper Coffee
--------	--------------------	-----------------

Both models, however, support the same answer:

Mary does the dishes before she drinks coffee.

The third sort of problems were “no-valid-answer” ones in which the premises support multiple models with no valid conclusion in common, e.g:

John has cleaned the house.

John has taken a shower.

John is reading the paper.

Mary always does the dishes when John cleans the house.

Mary always drinks coffee when John takes a shower.

What for Mary is the relation between doing the dishes and drinking coffee?

This problem is again consistent with multiple models:

Cleaning Dishes	Shower Coffee	Paper
--------------------	------------------	-------

and:

Shower Coffee	Cleaning Dishes	Paper
------------------	--------------------	-------

For this problem, there is no valid answer to the question, because the different models support a different relation between the two critical events (Mary drinking coffee and doing the dishes).

In general, the model theory predicts the following trend: one-model problems should be easier than multiple-model problems, which in turn should

be easier than no-valid-answer problems. This trend should be reflected in both the percentages of correct answers and the times taken to respond. Because the no-valid-answer problems call for a qualitatively different response, a critical comparison is between the one-model problems and the multiple-model problems, which both have valid answers.

Method

Subjects. 24 subjects carried out the experiment. They were all first-year psychology students, who were fulfilling a course requirement.

Design. The subjects acted as their own controls and carried out three kinds of problems, which were each based on five premises: (1) one-model problems with a valid answer, (2) multiple-model problems with a valid answer, and (3) no-valid-answer problems. Each subject carried out 12 inferences for each of the different sorts of problems presented with a different content, i.e. each subject solved a total of 36 problems. The problems were presented in a different random order to each subject.

Materials. The experiment was carried out in Leuven, Belgium and the materials were in Flemish (Dutch). Flemish distinguishes among the perfect, the present tense, and the aspectual description of future events in a very similar way to English, as shown by the following examples, which we have translated word for word:

John heeft een douche genomen.

(John has a shower taken.)

John neemt een douche.

(John takes a shower.)

John gaat een douche nemen.

(John goes a shower to take.)

Likewise, Flemish readily expresses temporal quantification:

Wanneer Mary de krant leest, neemt John altijd een douche.

(When Mary the paper reads, takes John always a shower.)

Each of the problems was based on a different set of such everyday activities selected at random from a pool of such events, and the names were selected at random from a pool of 36 female and 36 male first names.

The difference between the three sorts of problems depends in part on the first three premises. The one-model problems have a premise in the perfect, a premise in the present tense, and a premise referring to a future event; and so the order of the three events is fixed. The other two sorts of problems,

however, have a premise in the present tense, and either two premises in the perfect or two premises referring to the future. Hence, the order of the events is fixed, apart from the relation between the two events in the same tense. The multiple-model problems have fourth and fifth premises that refer back to the two events that have a fixed order; and the no-valid-answer problems have fourth and fifth premises that refer back to the events that do not have a fixed order.

We constructed 12 versions of the three sorts of problems in the following way. We used each of the six possible orders of the first three premises. The last two premises of each problem have to refer back to two of the events described earlier. The fourth premise can refer back to the earlier of the two events, and the fifth premise must then refer back to the later of the two, or vice versa. The combination of these two variables yields 12 versions of a problem. However, in the case of no-valid-answer problems, by definition, the two events have no fixed order, and so we used each of the six possible versions twice (with a different lexical content). As an illustration, Table 1 shows the 12 versions of problems in the multiple-model case.

Procedure

The subjects were tested in small groups of at most four subjects, and each subject was tested at a computer. At the beginning of each trial, the screen signalled "press space-bar for the next problem". When subjects pressed the space-bar, the first premise appeared and stayed on the screen until the subjects pressed the space-bar again. At this point, the next premise appeared, and the procedure continued until the fifth premise. When the subjects then

TABLE 1
The 12 Multiple-model Problems With a Valid Conclusion

<i>Problems</i>	<i>Premises</i>				
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
1.	past	past	present	while A	while C
2.	past	past	present	while B	while C
3.	past	present	past	while A	while B
4.	past	present	past	while B	while C
5.	present	past	past	while A	while B
6.	present	past	past	while A	while C
7.	future	future	present	while A	while C
8.	future	future	present	while B	while C
9.	future	present	future	while A	while B
10.	future	present	future	while B	while C
11.	present	future	future	while A	while B
12.	present	future	future	while A	while C

pressed the space-bar, the question appeared and the subjects typed their answer. The computer recorded six main latencies: the time taken to read each of the five premises, and the time from the presentation of the question until the subjects began to type their response.

The instructions were presented on the VDU. They explained that the subjects' task was to answer a series of questions based on information in previous assertions, and that the answers should be those that must be true given the truth of the previous assertions. If the subjects thought that there was no definite answer, they had to type that as their response. The subjects were told that although they would be timed, they should concentrate on making the correct responses. The instructions also explained how to use the space-bar to present the next premise and how to respond. The subjects solved one practice problem to make sure that they had grasped the procedure. None of the subjects had any difficulty with it.

Results and Discussion

The percentages of correct answers to the three sorts of problems were as follows: one-model problems: 85%, multiple-model problems: 78%, no-valid-answer problems: 58%.

This trend was highly reliable (Page's $L = 310.5$, $n = 24$, $P < 0.00005$), and the one-model problems were reliably easier than the multiple-model problems with valid answers (Wilcoxon's $T = 90$, $n = 14$, $P < 0.01$). Subjects tended to overlook the possibility of events occurring contemporaneously when they were both in the same tense: 79% of the errors to the problems with no valid answers overlooked this possibility, i.e. subjects answered that one event was happening before the other event, or that one event was happening after the other, and only 21% of the errors took the possibility into account (Wilcoxon's $T = 116$, $n = 15$, $P < 0.005$). The pattern of results clearly corroborates the predictions of the model theory.

Figure 1 presents the mean latencies for reading each of the five premises and for the response to the question. The data are shown separately for the three sorts of problems. There was a reliable trend in the total reading times for the five premises: subjects read one-model problems (23.6) more quickly than multiple-model problems (25.5), which in turn they read more quickly than no-valid-answer problems (25.5; Page's $L = 301$, $n = 24$, $P < 0.05$), and they read one-model problems reliably more quickly than multiple-model problems (Wilcoxon's $T = 221$, $n = 24$, $P < 0.05$).

The model theory makes a more detailed prediction about the reading times of individual premises: subjects should take longer to read any premise that leads to an indeterminacy in temporal relations and accordingly calls for the construction of multiple models. This comparison is easy to make in the case of temporal relations that are established using "before" and "after", but it is

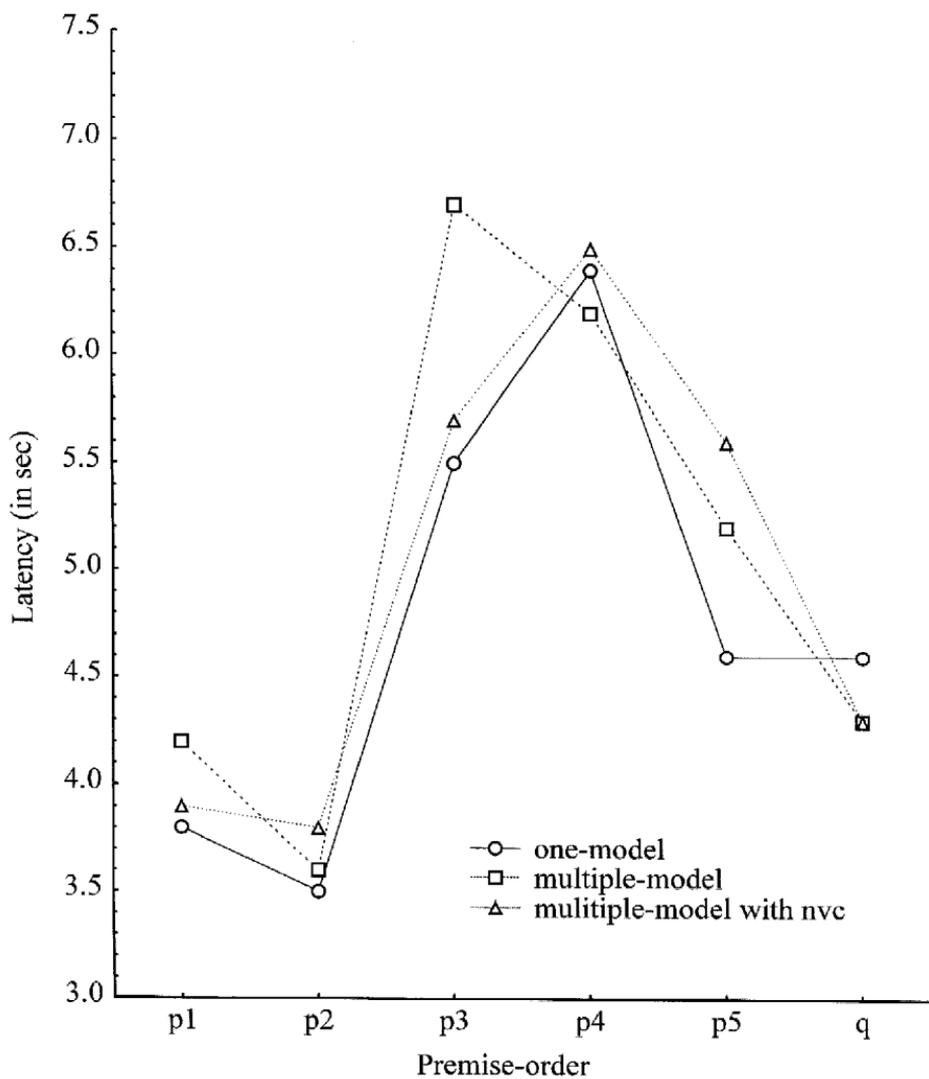


FIG. 1. The mean reading times in Experiment 1 for each of the five premises and the question for the three sorts of problems.

harder to make where the indeterminacy depends on tense and aspect. Consider, for example, the following sequence of assertions:

John has cleaned the house.

John is reading the paper.

John has taken a shower.

They establish that one of the events (reading the paper) occurs after the other two (cleaning the house, and taking a shower), but they do not establish the relative order of these two events. Hence, the third premise should take longer to read than the third premise of a one-model problem (this arrangement of premises occurs for eight out of the twelve versions of the problems, see Table 1). Now consider a different version of the problem:

John has cleaned the house.

John has taken a shower.

John is reading the paper.

Where two assertions in the present perfect occur one after the other, there is likely to be a natural tendency to treat the order of occurrence of the two events they describe as the same as the order of the two assertions themselves. The same tendency should occur when two premises referring to future events occur one after another. Hence, it may be more sensible to presume that even in these cases (the remaining four out of the twelve versions of the problems) reasoners will build an alternative model only after they read the third premise.

In fact, the reading times for individual premises provide somewhat mixed results. Unfortunately, there was a reliable difference in the reading times for the first premise of the three sorts of problems: 3.8 seconds for one-model problems, 4.2 seconds for multiple-model problems, and 3.9 seconds for no-valid-answer problems. The first premises of the one-model problems were read reliably more quickly than those of the other two sorts of problems (Wilcoxon's $T = 218$, $n = 24$, $P < 0.05$). As there is no way in which reasoners can distinguish the three sorts of problems solely on the basis of their first premises, and as the three sorts of temporal premises occur equally often in the three cases, this result can only be a consequence of a chance fluctuation. A re-analysis of the sum of the reading times excluding those for the first premise still yielded a reliable trend: the means were 19.8 seconds for one-model problems, 21.7 seconds for multiple-model problems, and 21.6 seconds for no-valid-answer problems (Page's $L = 307$, $n = 24$, $P < 0.005$), and the one-model problems were read reliably more quickly than the multiple-model problems (Wilcoxon's $T = 220$, $n = 24$, $P < 0.05$).

For the reading times for the second premise, no difference between the one-model problems and the multiple-model problems was expected, and there was no reliable difference (Wilcoxon's $T = 179$, $n = 24$, n.s.).

The third premise is the critical one for the multiple-model problems. Its reading times did not differ between multiple-model problems and problems with no valid answers, which both call for multiple models at this point (Wilcoxon's $T = 206$, $n = 24$, n.s.). However, as the theory predicts, these times were reliably longer than those for the third premise for the one-model problems (Wilcoxon's $T = 220$, $n = 24$, $P < 0.05$).

For the reading times for the fourth premise, there was no reliable difference between one-model problems and either the multiple-model problems or the no-valid-answer problems (Wilcoxon's $T = 150$, and 174 , $n = 24$, n.s.).

The fifth premise is critical for the problems with no valid answers, because subjects have to deal not only with two models, but also with the inconsistency between them. Hence, the reading times for the fifth premise should be the longest for these problems. Figure 1 shows the results, and the reading times for the one-model problems did not differ significantly from those for the multiple-model problems (Wilcoxon's $T = 175$, $n = 24$, n.s.). However, as the prediction suggests, the reading times for these two sorts of problem were reliably faster than those for the problems with no valid answers (Wilcoxon's $T = 210$, $n = 24$, $P < 0.05$).

Although the trend in the times to respond to the questions was in accordance with the model theory's predictions, it was not reliable (Page's $L = 277$, $n = 24$, n.s.). It seems that all the reasoning needed to answer the questions was carried out during the reading stage. This phenomenon, however, does not always occur (see e.g. Schaeken et al., in press).

Finally, and most importantly, there is the sum of the latencies for the third, fourth, and fifth premises and the question. Reasoners have to cope with one model while reading these premises in the case of a one-model problem, and they have to cope with multiple models in the case of multiple-model problems and problems with no valid answers. There was no difference between the sum of these latencies for the multiple-model problems (22.4) and the problems with no valid answers (22.1; Wilcoxon's $T = 149$, $n = 24$, n.s.). However, as the theory predicts, these times were reliably longer than those for the one-model problems (21.0; Wilcoxon's $T = 80$, $n = 24$, $P < 0.05$).

EXPERIMENT 2

Introduction

One problem with the previous experiment was the unexpected but reliable difference in the reading times of the first premise. A more serious problem, however, is a possible confound in the design of the experiment, which was pointed out by Nick Chater (personal communication). In the one-model problems, the first event is in the past and the third event is in the future. In the multiple-model problems, one event is in the present, and the other two events are in the past in half the problems, and in the future in the other half

of the problems. Hence, the temporal interval between the two events in the question is smaller for the multiple-model problems than for the one-model problems. It could therefore be harder to respond to the multiple-model problems, because the smaller interval may make for a greater confusion about temporal order. There is evidence both for and against this putative explanation.

The evidence for the explanation comes from a study carried out by Schaeken and Johnson-Laird (1995): in problems with six or eight events, the larger temporal interval yielded a more accurate performance than a small interval. The evidence against the explanation comes from an experiment reported by Schaeken et al. (in press). The subjects received two sorts of one-model problem describing the temporal order of five events in terms of the B-series, i.e. using "before" and "after". One sort of one-model problem had a smaller temporal interval between the two crucial events than the other sort of one-model problem. In particular; given three events in the order A B C, the first sort of problems concerned events contemporaneous with B and C, whereas the second sort of problems concerned events contemporaneous with A and C. However, as the results showed, there were no reliable differences in either accuracies or latencies between the two sorts of problem.

Given the equivocal nature of the existing evidence, we carried out a replication of Experiment 1 in which we made two crucial changes. The first change was in the tenses of the multiple-model problems, so that distance apart of the events in the questions was the same for all three sorts of problem: one-model, multiple-model, and no-valid-answer. In each case, one event in the question was in the past, and the other event in the question was in the future.

The second change was in the duration of the events in the sentences about the second protagonist, and therefore in the question itself. They now referred to momentary events, such as finding a key or dropping a cup, and in this way we minimised the possibility of subjects becoming confused by a possible overlap in time of the two events.

Here is an example of one of the resulting multiple-model problems (with a valid answer):

John has cleaned the house.

John has taken a shower.

John is going to read the paper.

Mary finds the key when John cleans the house.

Mary drops a cup when John reads the paper.

What for Mary is the relation between finding the key and dropping a cup?

Method

The design was the same as in the previous experiment. The lexical materials were the same, except that we used momentary events in the fourth and fifth premises. The one-model problems had a premise in the perfect, a premise in

the present tense, and a premise referring to a future event. The other two sorts of problems (multiple-model and no-valid-answer) either had one premise in the perfect and two premises in the future, or else one premise in the future and two premises in the perfect. The multiple-model problems had fourth and fifth premises that referred back to the two events with a fixed order; and the no-valid-answer problems had fourth and fifth premises that referred back to the two events that lack a fixed order. As before, we constructed 12 versions of the three sorts of problems. The procedure was the same as in the previous experiment, except that all subjects were tested individually. We tested 16 adult subjects, who participated voluntarily.

Results and Discussion

The percentages of correct answers to the three sorts of problems were as follows: one-model problems: 88%, multiple-model problems: 81%, no-valid-answer problems: 63%.

This trend was highly reliable (Page's $L = 219$, $n = 16$, $P < 0.001$), and the one-model problems were reliably easier than the multiple-model problems with valid answers (Wilcoxon's $T = 32$, $n = 8$, $P < 0.05$). This pattern of results clearly corroborates the predictions of the model theory.

Figure 2 presents the mean latencies for reading each of the five premises and for the responding to the question. The data are shown separately for the three sorts of problems. There was a reliable trend in the total reading times for the five premises: subjects read one-model problems (24.6) more quickly than multiple-model problems (26.2), which in turn they read more quickly than no-valid-answer problems (27.0; Page's $L = 202$, $n = 16$, $P < 0.05$), and they read one-model problems reliably more quickly than multiple-model problems (Wilcoxon's $T = 108$, $n = 16$, $P < 0.05$). Unlike the previous experiment, there was no reliable difference in the reading times for the first premise. There was no reliable difference in the reading times for the second premise. The reading times for the third premise did not differ between the multiple-model problems and the problems with no valid answers, which both call for multiple models at this point (Wilcoxon's $T = 88$, $n = 16$, n.s.). However, as the theory predicts, these times were reliably longer than those for the third premise for the one-model problems (Wilcoxon's $T = 95$, $n = 16$, $P < 0.05$). The reading times for the fourth premise did not differ reliably between the multiple-model problems and problems with no valid answers (Wilcoxon's $T = 72$, $n = 16$, n.s.). However, these times were reliably longer than those for the fourth premise for the one-model problems (Wilcoxon's $T = 104$, $n = 16$, $P < 0.05$). There was no difference in the reading times for the fifth premise. The trend in the times to respond to the questions was in accordance with the model and reliable (Page's $L = 224$, $n = 16$, $P < 0.001$). Finally, the sum of the latencies for the third, fourth, and fifth premises and the question did not differ between the multiple-model problems and the no-

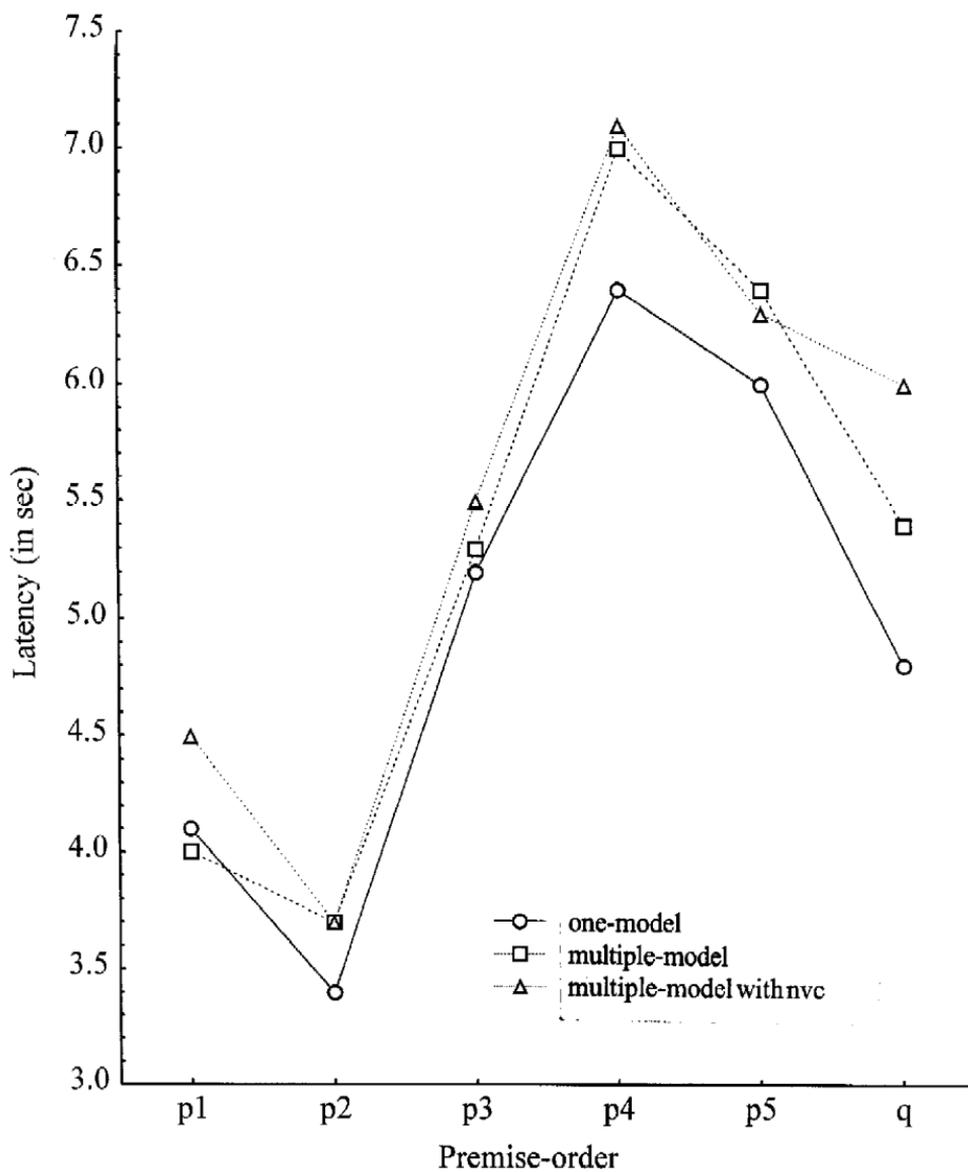


FIG. 2. The mean reading times in Experiment 2 for each of the five premises and the question for the three sorts of problems.

valid-answer problems (Wilcoxon's $T = 66$, $n = 16$, n.s.). However, as the theory predicts, these times were reliably longer than those for the one-model problems (Wilcoxon's $T = 108$, $n = 16$, $P < 0.05$).

GENERAL DISCUSSION

When subjects reason about temporal relations that are established by tense and aspect, they appear to do so by envisaging the relative order of the events. That is, they attempt to construct a mental model of this order. Where a description is determinate and accordingly calls for only a single model to be constructed, the task is relatively easy. However, it is reliably harder—as reflected by the percentage of correct responses—when reasoners need to envisage more than one model in order to do justice to the premises. It is hardest of all for those problems that lack a valid answer. In this case, a failure to consider the multiple models will lead subjects to make a definite answer where none is warranted. If more models have to be built, subjects also need more time. This phenomenon is clearly apparent in the times that the subjects took to read the premises. Moreover, those particular premises that call for constructing multiple models, or for grasping inconsistencies between models, took reliably longer to read, and thus presumably to interpret, than the corresponding premises in other problems that do not call for multiple models.

Both experiments, however, did result in some unexpected findings. In the first experiment, there was a reliable difference between the reading times of the first premises for the three sorts of problem. Because there are no linguistic differences between these premises, we infer that this result was a chance fluctuation. In the second experiment, the reading times of the fourth premise were reliably faster for the one-model problems than for the multiple-model and no-valid-answer problems. This difference may also be a chance fluctuation, but it could be a residual effect of the interpretation of the third premise, i.e. having to construct multiple models in the latter cases.

Although the sizes of the effects in both experiments are relatively small, the subjects probably constructed only two alternative models for the multiple-model problems, and so, unlike some other domains of reasoning (see Johnson-Laird & Byrne, 1991), the theory predicts only a relatively small difference. This consideration is particularly relevant because subjects who constructed only one of the two models for multiple-model problems with valid answers would still have produced the correct conclusion.

Is there an alternative explanation of our results? We counterbalanced the use of different temporal descriptions insofar as this procedure was compatible with the experimental design, and so it is hard to see how, for example, differences in the difficulty of interpreting sentences describing the past, the present, or the future, could explain our findings. Likewise, the sequence of such premises can vary in ways that seem intuitively more or less easy to

understand. However, again, these differences do not appear to be able to account for our results. The multiple-model problems with valid conclusions have an irrelevant premise, and in studies of spatial reasoning, Rips (1994) has argued that the presence of an irrelevant premise may make problems more difficult. Strictly speaking, however, the one-model problems in this experiment also have an irrelevant premise, i.e. the premise in the present tense, which is not referred to by any other premise. It might be argued that the position of the irrelevant premise is critical, i.e. whether it is the first premise, the second premise, or some other premise. In fact, the materials were designed to control systematically for the position of the irrelevant premise, which occurred equally often as the first, second, and third premise, in the one-model and multiple-model problems.

Could theories based on formal rules (see e.g. Braine, Reiser, & Romain, 1984; Rips, 1994) be somehow extended to deal with temporal reasoning of the present sort (the A-series of McTaggart, 1927, as we discussed in the Introduction)? In the case of B-series reasoning, which depends on “before”, “after”, and other such relations, it is a simple matter to introduce meaning postulates to capture their logical properties, e.g:

For any x, y, z , if x happens before y , and y happens before z , then x happens before z .

Schaeken et al. (in press) show in some detail how such a system might work although it is far from obvious how it could explain the difference between one-model and multiple-model problems. However, it is even harder to use formal rules to capture temporal reasoning based on tense and aspect (the A series). The required meaning postulates need to be very abstract and general, e.g:

If x is an event in the past, and y is an event in the future, then x happens before y .

If x is an event in the present, and y is an event in the future, then x happens before y .

A match between such assertions as:

John has cleaned the house.

John is taking a shower.

and these postulates is highly problematic if it is to be done in a purely formal way. And again, if this problem can be solved, it is difficult to see how the theory could account for the differences in difficulty between one-model and multiple-model problems. Ironically, systems of so-called “tense logic”, which have been developed by formal logicians, concern the A-series (see e.g. Prior,

1967; Rescher & Urquhart, 1971). These logics typically introduce operators that allow intervals of time to be specified, e.g. “for some moment(s) of time in the future, situation S is the case”. But, as a Reichenbachian analysis shows, the semantics of tense and aspect appears to be more complex than these logical systems (see e.g. Miller & Johnson-Laird, 1976, Sec. 6.2.4). Whatever the truth about these logics, however, there is no current psychological theory of temporal reasoning based on formal rules of inference—just as there is no such theory of modal reasoning in general.

Manuscript received 28 March 1995

Revised manuscript received 23 January 1996

REFERENCES

- Braine, M.D.S., Reiser, B.J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp.313–371). New York: Academic Press.
- Byrne, R.M.J., & Johnson-Laird, P.N. (1989). Spatial reasoning. *Journal of Memory and Language*, 28, 564–575.
- Fraisse, P. (1963). *The psychology of time*. New York: Harper & Row.
- Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, MA: Harvard University Press. Cambridge, UK: Cambridge University Press.
- Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- McTaggart, J.M.E. (1927). *The nature of existence*. Cambridge: Cambridge University Press.
- Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge, UK: Cambridge University Press. Cambridge, MA: Harvard University Press.
- Piaget, J. (1969). *The child's conception of time*. London: Routledge & Kegan Paul. [Originally published, 1927.]
- Prior, A.N. (1967). *Past, present and future*. Oxford: Clarendon Press.
- Reichenbach, H. (1947). *Elements of symbolic logic*. New York: Free Press.
- Rescher, N., & Urquhart, A. (1971). *Temporal logic*. New York: Springer-Verlag.
- Rips, L.J. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.
- Schaeken, W., & Johnson-Laird, P.N. (1995). How do subjects reason about temporal relations? In J.D. Moore & F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 725–730). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Schaeken, W., Johnson-Laird, P.N., & d'Ydewalle, G. (in press). Mental models and temporal reasoning. *Cognition*.