Title
How People Reason about Temporal Relations

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 17(0)

Authors
Schaeken, Walter
Johnson-Laird, Philip N.

Publication Date
1995

Peer reviewed

How People Reason about Temporal Relations

Walter Schaeken
Laboratory of Experimental Psychology
Department of Psychology
Tiensestraat 102
B-3000 Leuven
Belgium
walter.schaeken@psy.kuleuven.ac.be

Philip N. Johnson-Laird
Department of Psychology
Princeton University
Green Hall
Princeton, NJ 08544
USA
phil@clarity.princeton.edu

Abstract

The paper describes a theory of temporal reasoning and its implementation in a computer program. The theory postulates that individuals construct mental models, and it predicts that inferences that call for only one model to be constructed, such as:

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

will be easier than those that call for multiple models, such as a problem identical to the previous one except for its first premise: a happens before c. Experiment 1 showed that subjects were faster and more accurate with one-model problems than with multiple-model problems. They took more time to read a premise leading to multiple models than the corresponding premise in a one-model problem. Experiment 2 showed that if the question came first and was presented with all the premises, then subjects can ignore an irrelevant premise. As predicted, the difference between one-model and multiple-model problems with valid conclusions then disappeared. Experiment 3 showed that the size of a model, i.e., the number of events in it, and the distance apart of the critical events, also affected performance.

Introduction

Cognitive scientists have studied many aspects of time, but they have not hitherto investigated reasoning about temporal relations. Here is an example:

After the plane flew through the storm, the pilot radioed the tower.
The damage occurred while the plane flew through the storm.
What is the temporal relation between the damage and the pilot radioing the tower?

The answer, of course, is that the damage occurred before the pilot radioed the tower, and this answer is valid, i.e., it must be true given that the premises are true.

Temporal inferences are ubiquitous and often important, but how people make them is presently unknown. Our aim in this paper is to offer a solution to this problem. We first outline how the theory of mental models accounts for temporal reasoning, next we describe its implementation in a computer program, and finally we report the results of three experiments that corroborate the theory.

Mental Models and Temporal Reasoning: A Theory and an Algorithm

The theory of mental models postulates that reasoning -- deductive or inductive -- is a process in which reasoners first represent the meaning of premises, and then use this representation together with their knowledge to construct mental models of the relevant situations (see e.g., Johnson-Laird & Byrne, 1991). If a conclusion is true in all the models of the premises, then it is necessary (valid); if it is true in most of the models of the premises, then it is probable; if it is true in at least some model of the premises, then it is possible; and if it is true in only a few models of the premises, then it is improbable.

According to the model theory, the assertion:

After the plane flew through the storm, the pilot radioed the tower.

calls for a model that we represent in the following diagram:

```
s   r
```

in which the time axis runs from left to right, 's' denotes a model of the plane flying through the storm, and 'r' denotes a model of the pilot radioing the tower. Events can be conceived (and described) as momentary or as having durations, definite or indefinite. Hence, s is assumed to last for a certain duration that ends at some point prior to r. The assertion:

The damage occurred while the plane flew through the storm.

calls for the following addition to the model:

```
s   r
d
```

where 'd' denotes a model of the damage happening, and the vertical dimension in the diagram is used to represent contemporaneity. This model corresponds to infinitely many different situations that have in common only the truth of the two premises, e.g., the model contains no representation of the precise time at which the damage occurred. Yet, the conclusion:

The damage occurred before the pilot radioed the tower is true in this model, and it is not falsified by any model of the premises.

We have implemented a computer program (in Common Lisp) that carries out temporal inferences using models. The program has a context-free grammar for a fragment of English that contains assertions of the form, 'a happens before b', 'b happens while c', and so on. It constructs a semantic representation of any sentence in the fragment, using the meanings of words and semantic rules associated with each rule in the grammar. This semantic representation is then used to update the set of models. Given the semantic representation of the assertion: a happens before b, the program checks whether a or b is already represented in any model of the discourse. It then chooses the appropriate procedure:

1. If neither referent occurs in a model, the program starts a new model.
2. If one referent but not the other occurs in a model, the model is updated to include the new referent.
3. If the two referents occur in different models, these models are combined appropriately.
4. If both referents occur in the same model(s), then the premise is verified in the model(s). Because the program constructs all possible models of co-referential premises, verifying an assertion yields one of the following three responses: the assertion is a valid deduction (it is true in all the models); it was previously possibly false (it is false in some of the models, which are duly eliminated); or it is inconsistent with the previous premises (i.e., it is false in all the models). Given a question about the relation between two events, the program formulates a conclusion if a common relation holds between them 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 in any one model. For example, given the following problem:

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

the program constructs the following model:

```
a     b     c
d           e
```

from which it formulates the answer that d happens before e. Where the premises are not co-referential, as in:

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

each premise has a separate model, and the models cannot be integrated into a single model because they are not co-referential. Hence, there is no definite relation between a and d.

Some temporal descriptions are indeterminate, e.g.,:

a happens before c
b happens before c

do not fix the temporal order of a and b. In such cases, the program constructs models corresponding to the three possibilities: a happens before b, b happens before a, a and b happen contemporaneously. This procedure is, in principle, an intractable one: it yields an exponential growth in the number of models as the indeterminacies mount up. Yet, the procedure is feasible as long as there is only a small number of indeterminacies. Human performance degrades with an increasing number of models -- a predictable phenomenon if the human inferential system uses an intractable algorithm and a working memory of limited capacity.

The program is restricted in the number of models that it can construct in trying to solve a problem (by analogy with a limited capacity working memory). When the models it has constructed exceed this number, it then searches for a chain of premises interrelating the two events in the question, and constructs models only from them. (If there is no question, then the program cannot use this strategy.) Likewise, when human reasoners have immediate access to the question and the premises, they should construct models for just those premises that are relevant to the answer to the question. The advantages of this strategy are twofold. First, it ignores all irrelevant premises that are not part of the chain connecting one event in the question to the other. Second, it 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.

Three Experiments on Temporal Reasoning

The model theory predicts that an inference that depends on one model should yield fewer errors and take less time than one that depends on multiple models. We have carried out several studies to investigate this prediction, and Experiment 1 is such a study in which we also examined another prediction of the model theory: The time taken to read a premise that calls for the construction of multiple models should be longer than the time taken to read the corresponding premise in a one-model problem. This prediction cannot be made by a theory based on formal rules of inference -- if one were to be formulated for temporal reasoning -- because models would play no part in such a theory (cf. Braine & O'Brien, 1991; Rips, 1994).

We examined four sorts of deductions:

1. One-model problems with only relevant 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?

This problem has the following model:

a   b   c
d   e

If subjects were to use formal rules of inference, they would have to exploit the transitivity of 'before' to establish that a happens before c.

2. One-model problems with an irrelevant first premise (though this fact is not obvious to reasoners):

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

This problem has the following model:

a   b   c
     d   e

If subjects were to use formal rules of inference, this problem should be easier than the previous one, because there is no need to exploit transitivity, though they might be misled by the irrelevant first premise.

3. Multiple-model problems with a valid answer (but an irrelevant first premise):

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

This problem has at least the following two models:

a   b   c
     d   e

and

b   a   c
d   e

which support the answer: d happens before e. If subjects were to use formal rules, this problem should be no harder than the previous one because they have identical formal derivations. The second premise in problems of this sort calls for the construction of two alternative models, because of the indeterminacy between a and b. Hence, according to the model theory, it should take longer to read (and to interpret) this premise than the second premise of the one-model problems.

4. Multiple-model problems with no valid answer:

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

This problem has the following alternative models:

a   b   c
e   d

and

b   a   c
d   e

which do not support a valid answer to the question. Reasoners must consider both of these models in order to realize that there is no valid answer. In contrast, reasoners who overlook one of the models of a multiple-model problem with a valid conclusion may nevertheless draw the correct conclusion, which by definition holds for any model of the premises.

24 students at the University of Leuven with no training in logic acted as their own controls and carried out eight versions of each of the four sorts of problems, and the order of presentation was randomized for each subject. The materials concerned daily activities by two persons, the main clause was prior to the subordinate clause in each premise, and the eight different versions of each sort of problem were constructed in order to counterbalance the temporal connectives. The premises were presented one after another under the subjects' control, so that premise 1 disappeared when premise 2 appeared, and so on. The experiment was carried out by computer in order to record the reading time for each premise and the time to respond to the final question.

The subjects were correct for 93% of the one-model problems with only relevant premises, 89% of the one-model problems with an irrelevant premise, 81% of the multiple-model problems with valid answers, and 44% of the multiple-model problems with no valid answers. There was no reliable difference in accuracy of solutions between the two sorts of one-model problem (Wilcoxon's $T = 50$, $n = 11$, n.s.). However, there was the following reliable trend in correct answers: one-model problems were easier than multiple-model problems with valid answers, which were easier than multiple-model problems with no valid answers (by subjects, Page's $L = 315$, $n = 24$, $p < .0001$; and by materials, Page's $L = 111$, $n = 8$, $p < .001$). In addition, the one-model problems were answered correctly more often than multiple-model problems with valid answers (Wilcoxon's $T = 167$, $n = 20$, $p < .02$). The pattern of correct answers accordingly corroborates the model theory and is likely to run counter to any theory based on formal rules of inference.

The mean latencies to respond to the questions for the four sorts of problems were as follows:

1M problems with an irrelevant premise:	5.8 sec
Transitive 1M problems:	7.0 sec
MM problems with valid answers:	8.7 sec
MM problems with no valid answers:	10.7 sec

The response times to the question were faster for one model problems than for multiple-model problems ($F_{1,23} = 7.03$, $p < .02$). Likewise, the reading times of the second premise were reliably faster for one-model problems than for multiple-model problems ($F_{1,23} = 6.3$, $p < .02$). This difference corroborates the crucial prediction, because the second premise calls for the construction of alternative models in the multiple-model case, but not in the one-model case.

Multiple-model problems with a valid conclusion contain an irrelevant premise. Its presence is not the cause of their difficulty, because one-model problems with irrelevant premises are easier than the multiple-model problems. As our program shows, if the premises are interpreted under the guidance of the events referred to in the question, then irrelevant premises will be ignored and the multiple-model problems with valid conclusions thereby become one-model problems. In Experiment 2, we presented all the premises together either followed by the question or else preceded by the question. We tested 32 students at Leuven who acted as their own controls in a counterbalanced blocked design. When the question was last, the difference in difficulty between the one-model problems (94% correct) and the multiple-model problems with valid conclusions (86% correct) was reliable (Wilcoxon's $T = 90$, $n = 14$, $p < .001$). But, as we predicted, when the question was first, the difference in difficulty between the one-model problems (98% correct) and the multiple-model problems with valid conclusions (98% correct) disappeared. Multiple-model problems with no valid answer have no irrelevant premises and so should be unaffected by the position of the question, and indeed they remained difficult both when the question was first (56% correct) and when it was last (57% correct).

Multiple models are difficult because they increase the load on the processing capacity of working memory. The size of a model is likely to have a similar effect, and we have also carried out several experiments in order to investigate it. In Experiment 3, each sort of problem concerned either six events or eight events, and the temporal between the two events in the question was either small or large. An example of a 'small-interval' problem for one model with six events is:

a b c d
x y

and an example for one model with eight events is:

a b c d e f
x y

where x and y are the two events in the question. A previous experiment had shown that small-interval problems increase in difficulty as the number of events in a problem increases. In the 'large-interval' problems, the events in the question were always maximally apart, i.e., there were no irrelevant premises. An example of a large-interval problem for one model with six events is:

a b c d
x y

and an example for one model with eight events is:

a b c d e f
x y

A previous experiment had failed to detect any effect on accuracy of the size of the model with large-interval problems -- a surprising result if one believes in formal derivations since each extra event calls for an extra step in the derivation.

We tested 24 students, who acted as their own controls and carried out four versions of each of the different sorts of problems and 16 multiple-model problems with no valid answers. The problems were in a different order for each subject. The premises were presented one at a time on the computer screen under the subjects' control.

Table 1 presents the percentages of correct responses for each of the eight sorts of problems with valid conclusions. Overall, the one-model problems (86% correct) were easier than multiple-model problems (80% correct: Wilcoxon's $T = 136.5$, $n = 18$, $p < .05$). The large-interval problems were reliably easier

Table 1: The percentages of correct responses for each of eight different sorts of problems in Experiment 3.

	Small-interval problems	
	1M problems	MM problems
6 events	94	83
8 events	72	72
	Large-interval problems	
	1M problems	MM problems
6 events	95	92
8 events	83	73

than the small-interval problems (86% versus 80%; Wilcoxon's $T = 96$, $n = 15$, $p < .05$). The problems based on 6 events were significantly easier than the problems based on 8 events (77.4% versus 64.2%; Wilcoxon's $T = 282.5$, $n = 24$, $p < .0001$). The difference between the small-interval one-model problems based on six events and eight events was significant (Wilcoxon's $T = 78$, $n = 12$, $p < .005$). This difference was also significant for the large-interval one-model problems (Wilcoxon's $T = 91$, $n = 14$, $p < .05$). However, there was a significant interaction: The difference between the two sorts of small-interval problems was greater than the difference between the two sorts of large-interval problems (Wilcoxon's $T = 92$, $n = 15$, $p < .05$). Hence, although large-interval problems may increase in difficulty as the number of events increases, they do not do so as rapidly as small-interval problems. This factor may underlie the three-way interaction (Wilcoxon's $T = 31.5$, $n = 13$, $p < .05$): With 6 events, the difference in difficulty between one model and multiple models is evident in small-interval problems but not in large-interval problems (which may be showing a 'ceiling' effect); whereas with 8 events, the difference in difficulty between one model and multiple models is evident in large-interval problems but not in small-interval problems (which may be showing a 'floor' effect).

General Discussion

The experimental results suggest that reasoning about temporal relations depends on mental models rather than formal rules of inference. In general, the subjects in Experiment 1 read the four premises progressively faster, but contrary to this trend they took reliably longer to read a premise that led to multiple models than to read a corresponding premise in a one-model problem. Likewise, a description consistent with just one model yielded an easier inference than a description consistent with more than one model, and the difference held whether or not there was an irrelevant premise (and regardless of the length of a formal derivation of the conclusion). Vandierendonck and De Vooght (1994) have independently obtained similar results. We have also corroborated them in a study of temporal relations established by tense and aspect, e.g.,:

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.
What is the temporal relation between Mary doing the dishes and drinking her coffee?

Multiple models of temporal relations arise because of an indeterminacy, which in turn depends on an irrelevant premise for problems with valid answers. When reasoners can use the question to guide their processing of the premises, they can ignore the irrelevant premise and in this way convert the problem into a one-model one. Experiment 2 confirmed that subjects do indeed use this strategy, because the difference in difficulty between the two sorts of problems disappears in this case.

The first two experiments were based on the assumption that models of five events were small enough to be accommodated within working memory. As Experiment 3 showed, the size of a model does affect performance, and so too does the interval -- the number of intervening events -- between the two events in the question. Perhaps surprisingly, a larger interval yields a more accurate performance than a small interval. The cause of this effect is by no means certain. One possibility is that there is an 'end-anchor' effect, and another that the larger interval makes the order of events more discriminable (see Potts, 1978).

An earlier study of spatial reasoning showed that one-model problems were easier than multiple-model problems (Byrne & Johnson-Laird, 1989). Rips (1994, p. 415) has criticized this experiment on the grounds that the instructions biased the subjects "to use an imaginal strategy that favored the mental-model predictions (or placed a strong task demand on them to respond as if they were trying to image the arrays ...)". We are sceptical about these claims, but in any case neither the instructions nor the materials of the present experiments lend themselves to an imaginal strategy, real or simulated. In the absence of a rule theory for temporal reasoning -- and Rips does not advance one -- it is hard to rebut such theories, but we note that our results provide an interesting challenge to rule theories. They predict, for example, that one-model problems with an irrelevant premise have a shorter derivation than one-model problems with no irrelevant premises, yet Experiment 1 failed to detect any difference between them. If one argues, as Rips does, that the irrelevant premise led subjects astray, then one cannot explain why such problems are easier than the multiple-model problems with valid answers. They too have an irrelevant premise, and they have an identical derivation to the one-model problems with irrelevant premises. The irrelevant premise can play no part in a formal derivation, but it does play a part in the building of models: In the former case, the result is one model; in the latter case, the result is multiple models.

Finally, the use of mental models in temporal reasoning has one considerable theoretical advantage over other methods. It yields a decision procedure. An inference is valid if its conclusion holds in all the possible models of the premises, and it is invalid if it fails to hold in at least one of the possible models of the problems. Granted that problems remain within the capacity of working memory, then it is a simple matter to decide whether or not an inference is valid: One examines all the models of the premises.

Acknowledgements

Schaeken's research is supported by the National Fund for Scientific Research of Belgium and the Nuttin Foundation. Johnson-Laird's research is supported in part by the John S. McDonnell foundation.

References

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.

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

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Potts, G. R. (1978). The role of inference in memory for real and artificial information. In R. Revlin & R. E. Mayer (Eds.), *Human Reasoning* (pp. 139-161). Washington, DC: Winston.

Rips, L. J. (1994). *The Psychology of Proof: Deductive Reasoning in Human Thinking*. Cambridge, MA: MIT Press.

Vandierendonck, A., & De Vooght, G. (1994). Is reasoning with time concepts based on a spatialized representation of time? Report of the Department of General Psychology, University of Ghent, Belgium.